PHD

Bioprospecting for Extremophile Oleaginous Yeasts

Abd Ghaffar, Nur Rinah

Award date:
2017

Awarding institution:
University of Bath

BIOPROSPECTING FOR EXTREMOPHILE OLEAGINOUS YEASTS

Nur Rinah Abd. Ghaffar

A thesis submitted for the degree of Doctor of Philosophy
University of Bath
Department of Chemical Engineering
APRIL 2017

Acknowledgements

These four years that I have endured pursuing my dreams would never be realized without the smooth guidance of Dr Chris Chuck. He took a chance on me to go on board this project with my lacking in proficiency, but with some determination to fulfil her dreams. Thanks Chris for your patience, understanding and kindness throughout this journey. I am also humbled with the incessant help of Dr Daniel Henk, whom without I would not be able to finish my project. Thanks for letting me use lab 1.14 and the office room as if it was my second home. My sincere gratitude goes to Dr Fabio Santomauro who acts as my mentor from the very beginning of my endeavour. Thanks also to Professor Rod Scott, my co-supervisor, for allowing me to use the Department of Biology facilities though I was never even registered with the department.

To the lab gang: Tanakorn, Fraeya, Lisa, Rhod, Tristan, Joe, Jon, Rob, Felix, Sonia, Debs, Mauro and Kes, thanks for just being there and helping me out in many ways. Lab life would be just monotonous without you all tearing the lab down. It had been such great honour to get to know each and every one of you. I will not be able to repay the assistance of the technical team of the Department of Chemical Engineering, especially to Dan Lou-Hing, Fernando Acosta, Alexander Ciupa, Bob, Brigitte and others.

My life in Bath would not be the same without the craziness of my Malaysian undergraduate buddies. You are all full of life and make the word cheerful sounds boring. Thanks for taking me in as a sister for eternity. Thank you too, to all my long distance friends, who have been cheering me all the way to the finish line. You know who you are. Only god could reward all your prayers & wishes for me & my family.

From afar, I would like to show my enormous gratitude to my Papa, for without whom I would not even be here. I know you are always worried but I know too you have always believed in me. We know Mak (my mom) is looking over smiling from beyond. This is for you, Mak. Not forgetting my 2 siblings who have assisted me unrelentingly, especially on taking care of Papa.

My journey was not without heartaches & hurdles, and my two rocks are always there to tame the soul. Thanks my son, Raphaie, who has grown into a young man as I underwent this path, you have been so supportive and understanding. I could never repay the times that was stolen away from you, Ranya. I would always felt sorry. But your “It’s OK Mummy…” will always make it all alright. I am so blessed to have you both by my side.

My utmost gratitude be conveyed to MARA, my sponsor, an institution I was brought upon since the age of 16.

Lastly, I would like to once more thanks Chris for your believe in me. And everything would not materialized for me & my family without the grant from the Creator above.

Abstract

Palm Oil is the highest produced edible oil globally, with over 66 million tonnes produced annually. It has been estimated that up to 50% of all products sold in the supermarket contain palm oil in some form. Palm oil has attractive properties such as a high melting point and texture due to a balanced ratio of unsaturated and saturated fatty acids. It contains approximately 40% oleic acid (monounsaturated fatty acid), 10% linoleic acid (polyunsaturated fatty acid), 45% palmitic acid and 5% stearic acid (saturated fatty acid), that results in an edible oil that is suitable for use in a variety of food, detergent and cosmetics products. In addition, palm oil is the least expensive oil produced due to its high productivity and extensive production. Due to the high demand for the product, vast amounts of rainforest have been cleared to make way for more plantations, reducing biodiversity and releasing huge levels of carbon dioxide into the atmosphere.

There is a clear need for an alternative lipid that can match palm oils properties but can be produced sustainably. Recent work suggests that some yeasts are capable of producing a similar oil to palm oil and can be grown on waste resources. In this thesis a novel bioprospecting protocol was developed to isolate yeasts that can survive the harsh conditions necessary for industrial biotechnology. In this way a vineyard and the local area was sampled for yeasts which were then cultured under extremes of pH, multiple sugars and inhibitors caused from the breakdown of lignocellulose. The wild yeast were cultured in four stages: minimal medium with Lysine; minimal medium with inhibitors; minimal medium with xylose as sole carbon-source; and lastly minimal medium with only arabinose and cellobiose as carbon-sources. Only strains that survived each stage were taken forward to the next, to isolate species that were truly suited to these conditions. Out of the estimated 1000s of strains screened this resulted in 12 strains of yeast, mostly in the *Metschnikowia pulcherrima*, group being able to cope with the conditions.

The 12 strains were further analyzed by culturing them in an array of 4 different model lignocellulosic feedstocks namely wheat straw, corn Stover, sugarcane bagasse, and palm kernel cake hydrolysates. Other conditions incorporated in these analysis were a range of pH from pH 1.5 to pH 7.0; four levels of a mixture of 5 inhibitors; and two different temperatures. All of the 12 strains showed similar behaviour where inhibitor tolerance was only marked at higher pH, and at low pH the strains could not grow at all. Though all strains were able to grow on the hydrolysate models, even those with little glucose and/or xylose content. The lipid profile of the strains was also assessed and proved to be similar to most terrestrial crops, with suitable lipid profiles for a rapeseed oil, and in some cases palm oil substitute.

Lastly, to further evaluate the accurate identification of the strains as there are some ambiguity in the *Metschnikowia pulcherrima* group, we applied an approach only widely used for Pathogenic Bacteria/Yeast identification, Multilocus Sequence Typing (MLST). Using 25 strains (7 of this collection), 6 type species and some isolates from the original culture collection in Bath. Sequences of 6 genes was analysed using the Bayesian statistical method. The result showed grouping of *M. pulcherrima* into 3-4 groups 9 different for each gene. *M*. *Corniflorae* being the outgroup. In all 3 genes successfully sequenced: *M. Fruticola*; R6; Mp DAH 3; and ICS48 were consistently shown to be clonal.

The work presented here demonstrates a new method for bioprospecting strains capable of isolating strains for industrial biotechnology, and for characterisation of the yeast in the *Metschnikowia* genus. Some of the yeasts identified were oleaginous, and could potentially be used as a novel source of palm oil substitute.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Current vegetable oil use

Currently, vegetable oils and to a lesser extent animal fats are mainly used in the conventional food industry though further non-food uses, such as in the production of biofuels are increasingly impacting the sector. There are many characteristics that are vital in a food oil, these include high smoke point and low volatility for frying, in giving textures to the food (such as shortening, emulsion); providing flavour to bland oils (such as olive, sesame & almond oil), or as the carrier to the flavouring because, most chemicals that produce flavours and aroma are soluble in oil [1].

Worldwide, approximately 140 million tons of vegetable oils were produced in 2009/2010 and 70% of the production was accounted from only four sources: soybean oil, palm oil, rapeseed oil and sunflower oil [2]. In 2011/2012, about 25% of the world production of soybean oil (in USA, Argentina, Brasil and China that was 42.3 million tonnes) out of the world vegetable oil production [2]. Rapeseed oil accounts for 12% of the world production (22.1 million tonnes in 2010) and is mainly produced in Europe, China, India and Canada. Sunflower oil accounted for 13% of worldwide production (15 million tonnes) and is mainly sourced from European countries, Russia and Argentina. The majority of the world's palm oil production is centered in Malaysia, Indonesia and Africa, where production has increased from 24.3 million tonnes, in 2001 to 51.9 million tonnes in 2012. This makes palm oil the highest produced vegetable oil currently (Figure 1). Palm oil is also the most efficient producer with up to 4 t ha^{-1} y^{-1} produced from a typical South East Asian planation (Table 1.1).

Table 1.1: Productivity of Major oil Crops (2006/2007) adapted from reference [3]

Oil Crop	Oilseeds yield (t ha^{-1} yr^{-1})	Oil Content (%)	Calculated oil yield (t ha^{-1} yr^{-1})
Palm Oil (mesocarp)	19.03	20.1	3.82
Palm kernel	0.999	45.4	0.45
Cottonseed	1.28	14.7	0.19
Groundnut	1.04	43.2	0.45
Sunflower	1.25	41.2	0.52
Rapeseed	1.75	39.7	0.69
Coconut	0.52	66.1	0.34

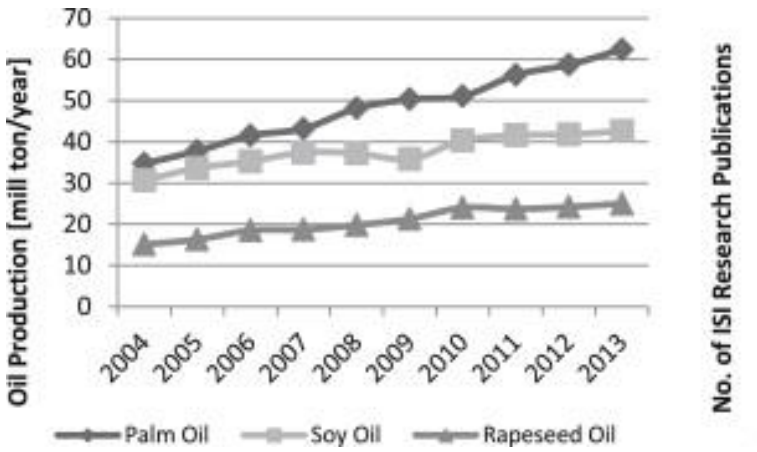


Figure 1.1: Oil production and research publications related to the three major vegetable oils [Oil production data from MPOB (2013)]

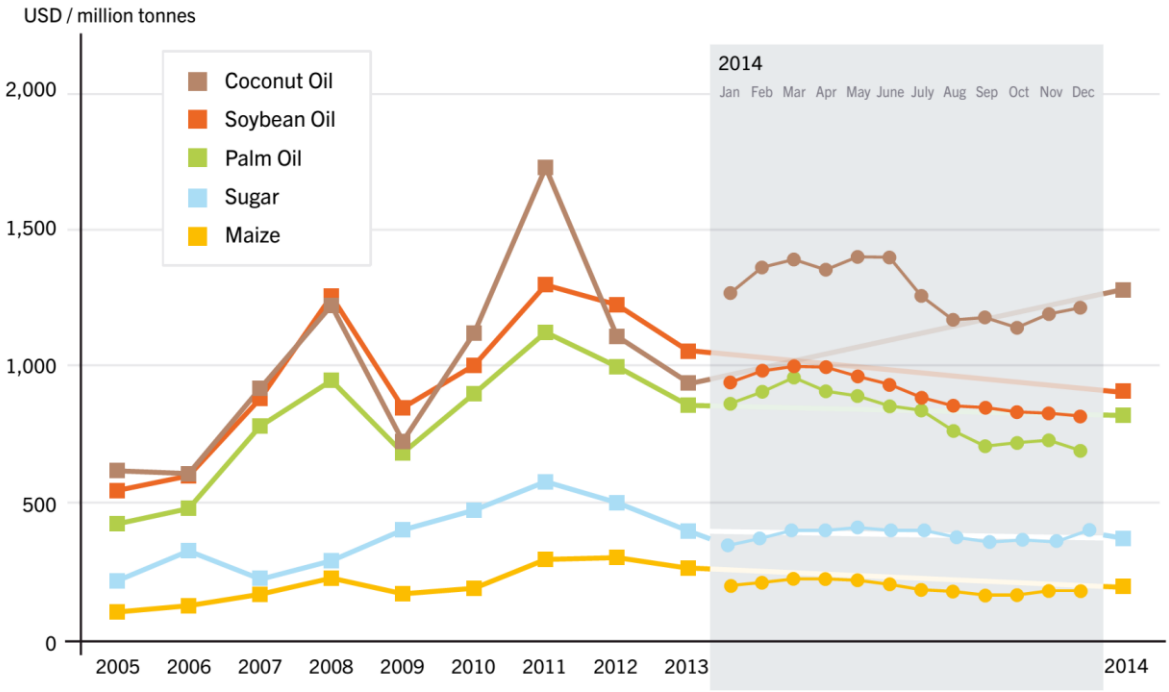


Figure 1.2: Market cost of vegetable oils over the period 2005-2014 (taken from Adib, 2015 [4])

1.2 Palm Oil

One of the main drivers for the rapid growth in palm oil production is the low cost. While the costs fluctuate, palm oil is always commonly known as cheaper than both rapeseed and soybean oil (figure 1.2).

Palm oil is an extremely versatile lipid product, with a wide range of uses in the food market. Palm oil itself can be fractionated, with the main palm oil product used for cooking, a mid-fraction that resembles cocoa butter so is extensively used in cosmetics, and palm stearin which has similar properties to tallow though is more cost-effective than these, as it requires less downstream processing stages for soap manufacture [5].

The main bulk of palm oil for food has two further important properties: a high melting point and a high degree of saturation. This makes palm oil ideal for cooking and gives the correct texture in the mouth when consumed. Other vegetable lipids, both naturally and processed, can have one of these two characteristics, but generally not both [5].

Technologically, it can be used as a major component of solid fats consistency without hydrogenation [6]. In Malaysia, refined palm oil requires only little reprocessing before use, mainly pressing & purification [7]. Other natural qualities that palm oil holds is that it imparts suitable crystallinity on blends and contains no trans double bonds [8].

Due to these factors, the demand for palm oil has risen dramatically and shows no sign of depreciating since the 1980's. This leads to a severe ecological problem. Palm oil plantations are situated in direct competition with tropical forest, which has led to the wide scale clearing, through burning of the indigenous forest land. In order to open new plantations, millions of hectares of forests were cut down. For example, approximately 7.5 MHa of rainforest was cleared between 1990 and 2010 on the Indonesian island of Sumatra [9-11].

While there has been a public backlash against the use of palm oil, especially with scientific studies demonstrating the catastrophic impact on the environment, this has not deterred the wider used in the food industry [12]. One of the main issues, is that other vegetable oils such as Soybean, have a lower yield per hectare and also similarly bad life cycle impact [13].

1.3 First generation biodiesel

An increasing use of palm oil is in the bioenergy section through the production of biodiesel. Biodiesel is the second largest produced biofuel currently with approximately 24 million tonnes produced each year worldwide [4]. This is produced almost entirely from the transesterification of vegetable oils such as rapeseed, soybean and palm [14].

However, over time a number of issues have become apparent with the use of biodiesel, particularly in long term storage. In these studies, the lubricity, acidity and water content increases over time, leading to excessive engine wear and premature failure. [15-17]. The higher viscosity of the biodiesel is another issue. The viscosity of the fuel is crucial in supplying sufficient lubrication for the engine parts but low enough to flow at operational temperature while higher viscosity may be sufficient lubrication for the engine parts but has to be low enough to flow at operational temperature [18]. The viscosity is related to the fatty acid profile, with more saturated esters giving a higher viscosity, as such palm oil derived biodiesel tends to have a high viscosity.

These issues have led to the development of new biofuel production techniques, such as hydrogenation to produce HVO (hydrogenated vegetable oil) [4]. 5 billion litres of HVO were produced globally in 2015, predominantly from palm oil. The hydrogenation does not require any specific fatty acid profile, though the lipids must have a high triglyceride purity.

As food crops are the main source of biodiesel and bioethanol, there is increasing concern over the environmental impact and negative carbon balance that the first generation technology has generated. This, coupled with the lack of suitable arable land for production, restricts the potential of these fuels [19]. The competition of the feedstock as a food source ultimately causes the increase in the feedstock prices because of the rise in the production of this fuel [20].

The available technology for such substitute can be exemplified from the biofuel technologies that have evolved becoming the alternative to fossil fuels [21] (figure 1.3).

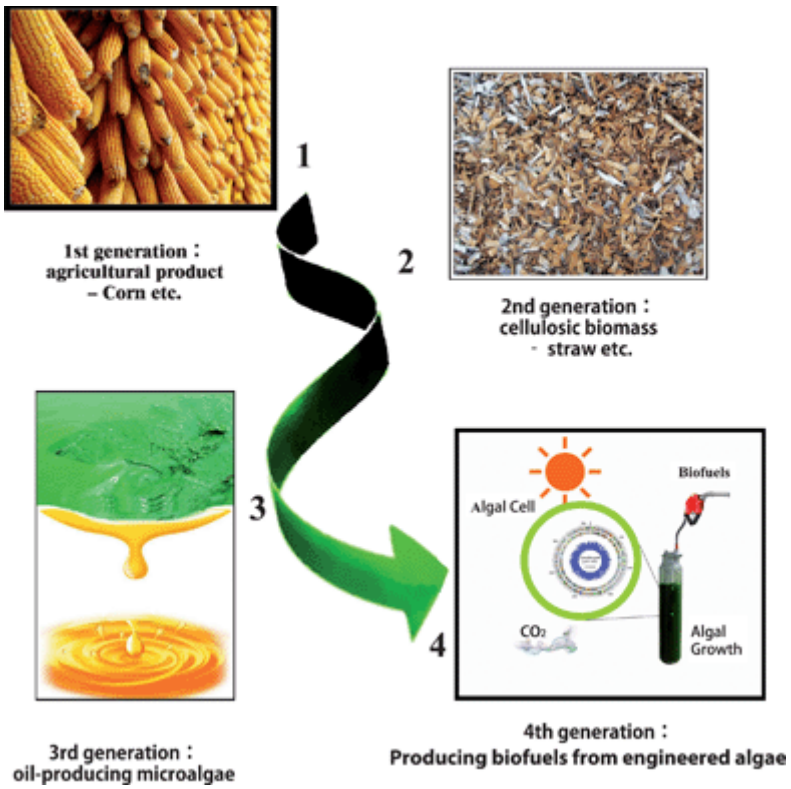


Figure 1.3: Four generations of biofuel production: agricultural products to algae [22].

1.4 Microbial conversion of lignocellulose

The impact of first generation biodiesel has led to a search for second generation alternatives. While waste cooking oil [23], coffee oil [24] and alternative plant oils such as jatropha [25] have all been suggested as possible replacements, there is only a small amount available for conversion and the sustainability is questionable. Seemingly to produce a more sustainable source of triglyceride oils, then lignocellulose as a feedstock must be accessed. One possibility to accomplish this is through culturing oleaginous yeast capable of growing on the hydrolysate produced [26].

To date, only bioethanol production has been commercialised as a method of converting lignocellulosic biomass through fermentation. Similarly to biodiesel, the main production worldwide of bioethanol is still from the fermentation of first generation feedstocks. This is mainly sugarcane derived sucrose in Brazil or corn derived starch carbohydrate in the USA [27]. To produce sustainable fuels that do not impact on food production, the industry has turned to using cellulose as a source of sugars for

bioethanol production (fig 1.4). Cellulose can be produced from non-edible wastes, as well as non-crop feedstocks such as trees and grasses, forest residues (sawdust), industry residues (black liquor from the paper industry), agricultural residues (corn stover), municipal waste and sustainable biomass (jatropha, camelina and switchgrass) [28, 29].

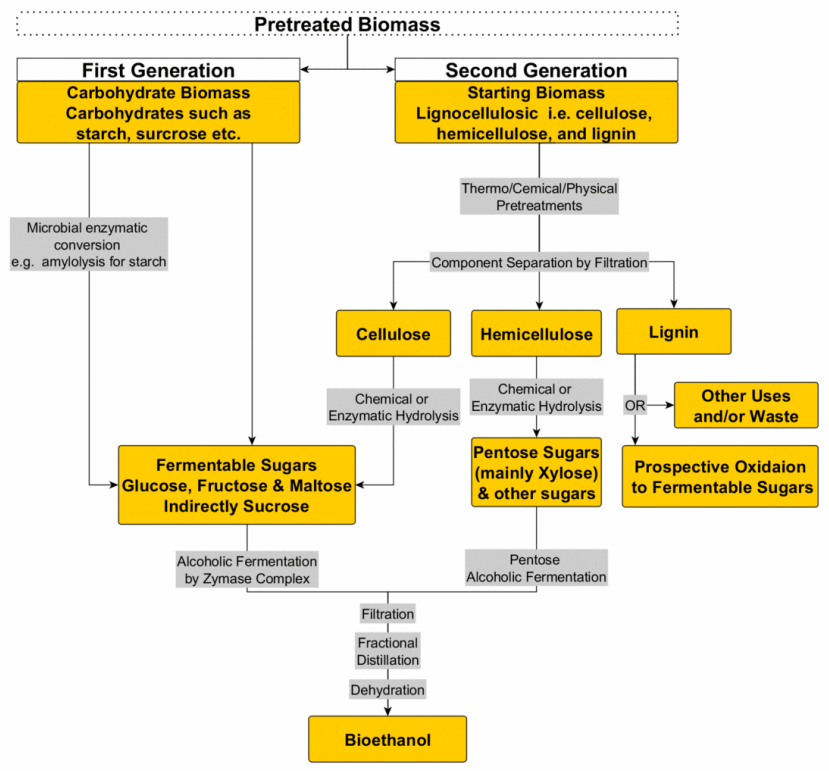


Figure 1.4: A simplified flowchart of the 1st and 2nd Generation Biofuel processes (https://creativecommons.org/licenses/by-sa/3.0/deed.en)

Lignocellulose is the most abundantly available raw material on the Earth. On a worldwide level, it is estimated that potentially 1,500 million tonnes of lignocellulose are available for fuel production. In Europe, straw production is estimated to yield approximately 120 million tonnes of fermentable material [30], while in Asia, 700 million tonnes, of rice stover predominantly, are produced [31]. However, alternative studies have suggested there is even more biomass available. In the U.S.A. alone, analysis done circa 2006 showed that roughly 1.3 billion dry tonnes of biomass each year could be produced alongside existing agricultural and forestry production [32].

The production of fuels from lignocellulose entails a huge number of steps from feedstock selection to process engineering, all of which contain challenges. Due to the

complexity of the lignocellulose material (LCM) these feedstock require advanced processing mechanisms which are expensive and on a scale that has not been realised out of the refinery industry [33].

Lignocellulose waste is the most abundant, secured and renewable feedstock available for the current usage and feasible development of bio-product [34]. Second generation lignocellulose waste such as agricultural stover, forestry waste, or dedicated energy crops such as switchgrass, exhibit a lower carbon debt for land use compared to other first generation sources, besides eliminating the competition between food and fuel [35-37]. Their advantageous characteristics give them great promise for biotechnological applications [38]. It has been estimated that, globally, 5.2 billion tonnes of biomass can be available for less than $60 per tonne by 2030, much of this derived from agricultural waste from the 2.3 billion tonnes of grain produced worldwide in 2011 [39, 40].

At present, cultivation wastes are possibly burned, turned over into the soil, used as animal feed or discarded into landfills. Due to short crop cycles, rice straw is regularly burnt in the fields, resulting in air pollution hazardous to human health [41]. However, the huge amounts of lignocellulosic biomass that are available on the planet can potentially be converted into a variety of different value-added products [42, 43] as depicted in figure 1.5. This includes bio-fuels, carbon sources for microbial fermentation and their enzyme production, chemical productions, pulp and paper production, animal feedstock, and polymer for composite materials.

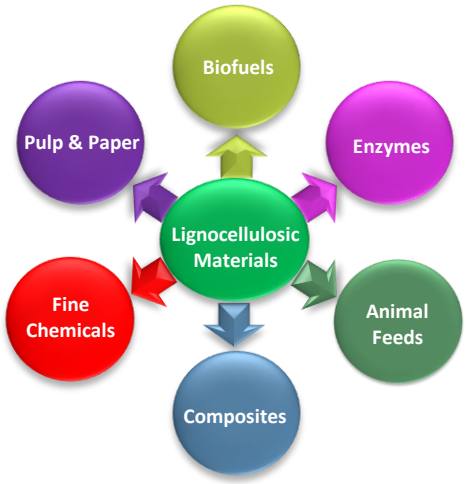
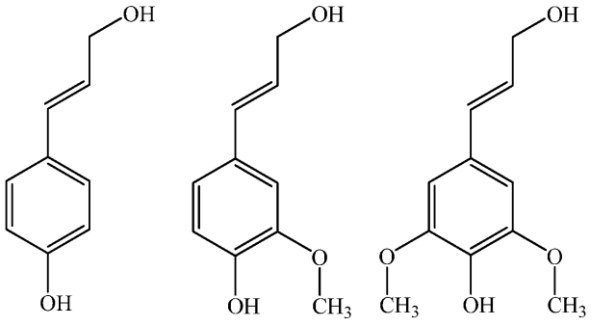


Figure 1.5: Bioconversion of Biomass into value-added products (adapted from Iqbal et al. [44]).

1.5 Lignocellulose Depolymerisation

Lignocellulose are structural materials that are produced by plant cells to construct cell walls, leaves, stems, and stalks are constituted primarily of three different types of biochemical, categorized as cellulose, hemicellulose, and lignin [45].

Lignin is a copious polymer in the plant cell walls, giving the structure firmness and as a result protects cellulose polymer from the hydrolytic attack by pathogens. Chemically, it has been described as hydrophobic three-dimensional phenylpropanoid precursor derived from the random coupling of lignin monomers; coniferyl, syringyl and p-coumaryl alcohols. Due to its complexity in molecular texture, it is highly resistant to chemical and microbial degradation [46].



coumaryl alcohol coniferyl alcohol sinapyl alcohol

Figure 1.6: Lignin monomers (adapted from Ralph et al. 2017 [47])

Hemicellulose is a heterogeneous group of compounds that in plant-cell walls form part of the matrix within which cellulose fibres are embedded, that are characterized by being neither cellulose nor pectin. The two principal structural types are polymers of D-xylose and of glucose and mannose, known respectively as xylans and glucomannans [48]. Hemicellulose is frequently branched with degree of polymerization of 100 to 200 [49].

Cellulose, a polymer of glucose, consists of linear chains of (1,4)-D-glucopyranose units, each units are linked 1–4 in the *β*-configuration, with an average molecular weight of around 100,000, which is very difficult to hydrolyse under natural conditions [50]. The degree of polymerization of cellulose chains range from 500 to 25 000 [49]. It is the major chemical component of fiber wall and is contributing 40-45% of the wood's dry weight [38].

Table 1.2 Compositional analysis of representative common lignocellulosic feedstocks. [51]

Lignocellulosic materials	Carbohydrate composition (% dry wt.)		
	Lignin	Cellulose	Hemicellulose
Agricultural residues	5–15	37–50	25–50
Banana waste	14	13.2	14.8
Bagasse	23.33	54.87	16.52
Barley hull	19	34	36
Barley straw	6.3–9.8	36–43	24–33
Bamboo	23	49–50	18–20
Corn straw	8.2	42.6	21.3
Corn cobs	15	45	35
Corn stover	19	38	26
Cotton seed hairs	0	85–95	5–20
Coffee pulp	15.6–19.1	33.7–36.9	44.2–47.5
Douglas fir	15–21	35–48	20–22
Eucalyptus	29	45–51	11–18
Grasses	10–30	25–40	25–50
Horticultural waste	36	34.5	28.6
Hardwood	18–25	40–55	24–40
Olive tree biomass	19.1	25.2	15.8
Jute fibers	21–26	45–53	18–21
Leaves	0	15–20	80–85
Nut shells	30–40	25–30	25–30
Newspaper	18–30	40–55	25–40
Oilseed rape	14.2	27.3	20.5
Oat straw	10–15	31–35	20–26
Poplar wood	10–21	45–51	25–28
Pulp and paper sludge	16	23.4	8.6
Pine	23–29	42–49	13–25
Rice Straw	18	32.1	24
Rice husk	15.4–20	28.7–35.6	11.96–29.3
Sugar cane bagasse	20	42	25
Sweet sorghum	21	45	27
Softwood	25–35	45–50	25–35
Sponge gourd fibers	15.46	66.59	17.44
Sorted refuse	20	60	20
Solid cattle manure	2.7–5.7	1.6–4.7	1.4–3.3
Swine waste	NA	6	28
Sugar beet	NA	5	5.5
Winter rye	16.1	29–30	22–26
Wheat straw	16–21	29–35	26–32
Water-hyacinth	3.55	18.4	49.2
Wheat bran	8.3–12.5	10.5–14.8	35.5–39.2

Arguably, the greatest economic barrier and challenge is to cost-effectively depolymerise the lignocellulose. There are two main methods to convert lignocellulose into biofuels and bio-products: thermochemical and biochemical conversions [33]. The thermochemical conversion process is shorter but requires more energy input, while the biochemical conversion, in theory, has higher returns with low energy consumption due to the moderate reaction conditions [52]. In practice a mixture of the two processes are used.

Conversion of lignocellulosic biomass into bioethanol consists of the following steps [53]:

(i) Mechanical treatment of the lignocellulose to lessen recalcitrance in the lignocellulose

(ii) Chemical pretreatment of the lignocellulose to break down and solubilise the polysaccharides into oligosaccharides

(iii) Hydrolysis of the solubilised hemicellulose and cellulose by enzymes

(iv) Fermentation of the saccharides to ethanol and other metabolites by microbes

(v) Distillation to purify the fuel

The original mechanical and chemical pretreatment stages of the lignocellulose is basically to open the compact structure so that substrates can penetrate better, therefore maximizing the amount of cellulases that come in contact with the cellulose [54].

Pretreatment of wood-derived lignocellulosic is essential for ensuing an effective fermentation to ethanol, as wood tends to be more stubborn than normal crop plants because of the higher lignin content, the intricate ultrastructure and the increased difficulty in hydrolysis of the structural polysaccharides [55]. The pretreatment process breaks up the structural walls, by inducing the formation of pores that improve enzymatic accessibility [56].

The physical pretreatment stage includes milling, extrusion or microwave treatments. They are usually combined to produce the best results. In each case, the density of the

lignocellulose is being taken into consideration. Milling, though will produce optimum chip sizes, is the most inefficient in energy usage. Extrusion, a thermo-physical treatment, is costly but does not generate inhibitors such as furfural and HMF. Microwave technology has a short processing time while using less energy, though has not been proven on the industrial scale required for fuel production [57].

In addition to the mechanical, a range of chemical pretreatments have been developed [58]. The main chemical pretreatments used are:

i) Autohydrolysis; where the cellulose is heated at high temperatures and pressures with water to aid break down

ii) Ammonia fiber explosion (AFEX); where ammonia is used to aid the breakdown process

iii) CO_2 explosion; the cellulose is broken down by the sudden release of high pressure CO_2

iv) Alkaline hydrolysis; alkali catalysts such as NaOH can be added to aid the hydrolysis reaction

v) Acid hydrolysis; Strong acids, such as sulfuric and nitric, or organics such as acetic acid are used to depolymerise the cellulose polysaccharide

vi) Ozonolysis; can be used to oxidatively cleave the lignin and saccharides and aid solubilisation.

viii) Enzymatic treatments, a less harsh pretreatment is the biological treatment of the lignocellulosic material. Two lignolytic enzymes, namely phenol oxidase (laccase) and peroxidases (lignin peroxidase, LiP and manganese peroxidase, MnP) from white and soft-rot fungi are utilized to help breadown the lignin structure and hence the macro LCM structure [59, 60].

After pretreatment, various cellulases can be used to produce the monosaccharide and disaccaharides that can be readily fermented to fuel molecules. Cellulase enzymes carries out enzymatic hydrolysis and are highly specific in production of the reducing sugars including glucose [58]. Overall cellulosic hydrolysis involves at least three key enzymes: an endoglucanase, an exoglucanase, and a β –glucosidase [27]. Certain

organisms have the ability to hydrolyse by cellulosomes, cell-associated extracellular multienzyme macromolecule, binding cellulases and hemicellulases, which is very efficient in producing sugars [61]. Certain types of fungus, such as *Trichoderma* sp. that have been reported to produce cellulases naturally and can be ulitilised efficiently [58]. The main sugars produced from this process are glucose, xylose, cellobiose and arabinose.

1.6 Inhibitory substrates

As lignocellulose goes through the aggregated pretreatments, the sugars produced can also break down further into inhibitory by-products. The pretreatment usually break down the hemicellulose into pentose and hexose sugars, these can then further break down into sugar acids, acetic acid, formic acid, levulinic acid, furan aldehydes, 5-hydroxymethylfurfural (HMF) and furfural [62]. After hydrolysis of lignocellulose polysaccharides, lignin generally remains intact as a solid, though a small amount is degraded to phenolics and other aromatic compounds under especially harsh conditions [63].

The inhibitory compounds are divided into three categories based on their source as above. They are weak acids, furan derivatives, and phenolic compounds [64]. All the inhibitors can hinder the biological processes in different parts of the process. For example, undissociated weak acids are liposoluble and dispersed across the plasma membrane, the low dissociation of H^+ into the cytosol inhibits growth of yeasts and bacteria [65]. The cell reproduction also decreases as the pH decreases [66]. These acids have been shown to severely reduce ethanol production in *S. cerevisiae* [67].

One theory on the negative impact of acids on cellular growth is that the ATP hydrolysis is at a high, hence the proton-pumping capacity is exhausted by the cell (to maintain the intracellular pH) [66]. Alternatively, the anion accumulation theory has also been put forward. In this theory high anion accumulation in the cell will create internal acidification and directly interfere with the cell, inhibiting growth [68].

Furfural compounds are also inhibitory. During fermentation yeasts convert furfural to furfuryl alcohol, this has a similar effect on cell growth as ethanol [64]. Interestingly, for

S. cerevisiae, glycerol production was proven to be significantly lower during furfural reduction, suggesting that furfural reduction regenerates NAD^+. Reductions of furfural to furfuryl alcohol and acetaldehyde to ethanol were presumably using the same enzyme, alcohol dehydrogenase (ADH) as suggested by Palmqvist et. al. [68], this potentially reduces ethanol production through competition. Similarly, HMF is converted to 5-hydroxymethyl furfuryl alcohol, but at a much slower rate than furfural. It is suggested that it goes through the same mechanism of inhibition [65].

1.7 Oleaginous yeasts

A small number of yeasts are well known to be oleaginous, and produce over 20% dry weight in triglycerides, similar to the main components of plant and animal lipids. These lipids can be used as a feedstock for a range of industrial applications.

The yeast lipids are mainly composed of triglycerides, though contain varying levels of free fatty acids, other neutral lipids such as mono- and diacylglycerides and sterol-esters, sterols and polar lipids e.g. phospholipids, sphingolipids, glycolipids [69].

Yeast oils are particularly advantageous and offer many advantages over vegetable oils including a far higher production yield per year, less labour intensive, and will be less affected by location, season and climate change. The yeast *Yarrowia lipolytica* is arguably the most researched organism of this class, though out of over 1600 known yeast species, at least 40 are thought to be oleaginous. These yeasts come from the genera *Candida, Cyberlindnera, Geotrichum, Kodamaea, Lipomyces, Magnusiomyces, Metschnikowia, Trigonopsis, Wickerhamomyces, Yarrowia, Crypotcoccus, Guehomyces, Leucosporidiella, Pseudozyma, Rhodosporidium, Rhodotorula* and *Trichosporon* [26, 70].

Sugars produced from depolymerisation of lignocellulose are varied though are generally a mixture of hexose and pentose sugars, predominantly these are glucose and xylose that are present in a ratio of 2:1 [71]. A significant amount of the cost is derived from the cost of the initial feedstock, and therefore it is vital that all the sugars are used on depolymerisation, including low level ones such arabinose or cellobiose.

While there has been little economic analysis undertaken on the production of biofuels from yeast oils, the two largest factors are the cost of the initial feedstock and the

fermentation. To reduce these costs and deliver a more economic product, then only the most inexpensive feedstocks can be used and less costly fermentation routes must be considered. To this end yeast that can grow on a range of sugars, with high inhibitors (consisting of acids, furans and phenolics) will allow lower production costs of the lignocellulose, while pH tolerance would make a more stable system, less likely to be invaded by other species [72, 73].

1.8 Bioprospecting methods for yeasts suitable for biotechnology

'Biodiversity prospecting' or Bioprospecting is defined as the discovery of biodiversity for beneficial and marketable, biochemical and genetic reserves to attain commercial advantages or for conservation purposes [74]. It pertains to plants, animals and all living organisms, including microbes, such as bacteria and fungi. Bioprospecting can also include the compilation of traditional knowledge concerning to the use of these methods from local populations [75].

Pressing environmental problems and biotechnological advancements have led to the changes in numerous chemical industries where many chemical catalysts are being substituted with suitable biocatalysts such as enzymes [76]. Microorganisms offer an excellent well of enzymes and bioactive compounds that can be used as biocatalysts, bacteria and yeast are of special interest due to their ability to be grown in dense cultures over a shorter period, and can be produced generously and consistently [77]. For example, using a bioprospecting methodology a range of novel cellulases have been discovered and further developed for the depolymerisation of lignocellulose. This concept could potentially be used to find suitable platform organisms for the production of biofuel precursors [78].

While scientists have generated a myriad of chemical compounds which are not generally found in nature, such as the synthetic drug Valium for example, natural compounds are becoming increasingly important in the pharmaceutical industry for drug discovery. From a list of the thirty most popular pharmaceuticals in 1997, thirteen were developed from naturally derived compounds isolated from wild plants, animals, and microorganisms [79]. Natural products have been used as drugs without further

chemical conversion or form a platform for a range of drugs that are then synthesized chemically. As of 2002, there are over 200,000 natural metabolites recorded with some bioactive properties, of interest to the pharmaceutical market. For example, more than half of the new chemicals launched into the market worldwide between 1981 and 2002 were natural products or their derivatives [80].

Two recent examples of successful drugs were isolated using a bioprospecting approach. The first is a range of anti-obesity drugs produced by the UK Company, Phytopharm. These are based on a series of related plants found in the Kalahari called ghaap/xhooba, or *Hoodia gordonii* [81]. Another example is the anti-cancer drug taxol, which was originally derived from the bark of the Pacific Yew tree but is now produced using tissue culture techniques [82].

1.8.1 The normal route of Bioprospecting.

According to Mateo, et al. [79], Bioprospecting can be divided into these three types (figure 1.6), these are:

1. Chemical Prospecting
2. Gene Prospecting
3. Bionic/Mechanic Prospecting

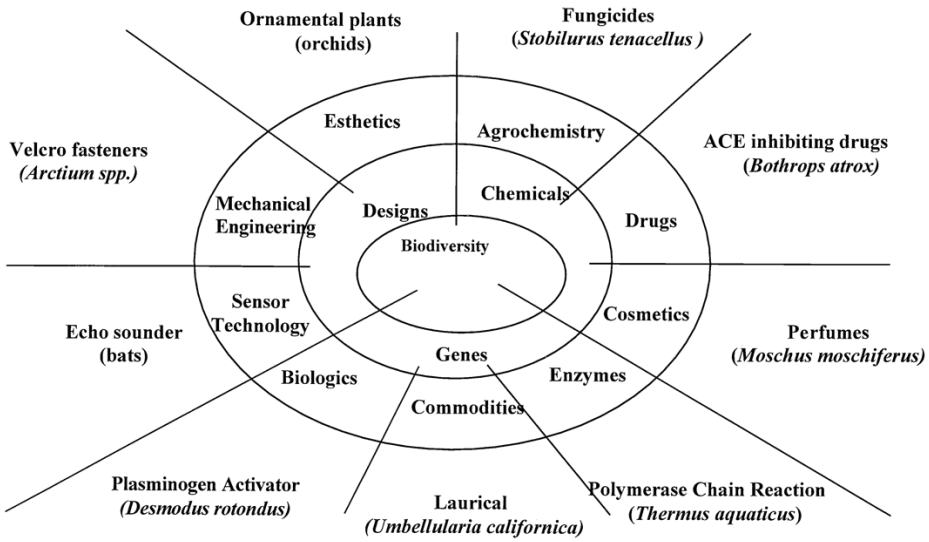


Figure 1.7: The three principle sources of Biodiversity their applications. (taken from Mateo et. al. [79])

In the finding of novel natural products from Biodiversity, the majority of work has been through looking at the chemical output (Chemical Bioprospecting). Where the impact of the organism and its metabolites on the environment are used as the key indicator. This is true for the examples given in the previous section. However, the chemistry concept here is quite broad in range, the development of diversity in the extensive scale of chemistry can be attributed to the competition, communication, sexual attraction, and pollination at both species and molecular level [83].

Genetic prospecting targets the phenotypes (thus, genotypes) from the wild type flora and fauna. A very significant example was the discovery of the protein, Desmodus Plasminogen Activator (DPA), that dissolves thrombolytic blood clots for vampire bats (*Desmodus rotundus*) to drink blood clot-free [84]. Recently the recombinant human tissue Plasminogen Activator (tPA) has been approved by FDA and for use in Europe as a potent therapeutic method for reviving heart attacks [85], entirely based on this work.

A further method of bioprospecting is Bionic prospecting. This is used generally by structural engineers and architects to exemplify the nature for construction and technical solution. Velcro fastener would be the most commonly used product that inventors got the inspiration from the tackiness of the seeds of a common weed, Burdock (*Articum spp*.) [86]. Recently this definition has been expanded, into studying the whole system of a microbial community or an insect community and their interactions, for example on butterflies and grasshoppers in Costa Rica. The new coupling mechanisms led to the original branch in engineering of “tropical bionic,” where scientists exemplify how these eukaryotes of an array of species adapt to their ecological niche [79].

Due to the vast potential of bioprospecting, the United Nations and many developing countries with abundant natural habitats and forest, have drawn up substantial guidelines for the management of these activities [87]. The Convention on Biodiversity (CBD) was created to have biodiversity protecting in all aspects including in the application of Bioprospecting [88].

In the last few years, with increasing biotechnological advancements, bioprospecting is starting to be used to generate a wealth of new organisms, genes and other components

to produce industrial products. Bioprospecting ultimately goes hand-in-hand with these biotechnological advances. As such, more stringent policies and laws are starting to be enacted to guide the scientific exploration and uphold the sound application of Bioprospecting [89]. Table 1.3 lays out the potential synergies and policies that would be of future use in both Bioprospecting and biotechnology.

Table 1.3: Potential synergies between biotechnology development and value-added bioprospecting [89]

Scientific and technical linkages	Common policy and programmatic issues	Reciprocal benefits
Biochemistry	Foreign investment policies	Creation of conservation incentives
Genetics	Technology licensing arrangements	Additional sources of funding and technical assistance
Cell & tissue culture	Intellectual property rights to isolated biochemicals	Broader allocation of policy and program development costs
Fermentation techniques	Coordination of public–private R&D activities	Diversification of market opportunities
Recombinant production of natural products	Finance and business development for start-up enterprises	
Prospecting for genes conferring valuable agronomic traits or coding for valuable enzymes and other products	Access and benefit sharing for use of wild biodiversity	

1.8.2 Bioprospecting microorganisms for biotechnology application

One of the first targeted bioprospecting approaches for lipid production, was not in using oleaginous yeasts, but rather investigating microalgae for their potential to produce fuel suitable lipids.

A range of techniques have been applied on a suitable Bioprospecting method for microalgae, which has been well reviewed by Mutanda and co-workers (2011) [78]. In this review, the authors focussed on the specific aims of these bioprospecting studies, which targeted identifying the algal species with outstanding lipid production, focusing on organisms with high growth rate all of which were acquired from different habitats.

The microalgae investigated were autotrophic organisms which were grown photosynthetically, in the study the largest abundance were diatoms, green algae, blue-

green algae and golden algae [90]. The sampling considered temporal and spatial effects for a successful collection to be gained. As this Bioprospecting regime was designed for the production of biofuel, and microalgae have been known to thrive in many extreme environments, the habitats selection was imperative, and a range of parameters were investigated at the collection site.

Figure 1.8 shows a flow diagram of the Bioprospecting system used. After acquiring the samples further steps were taken to isolate the strains. Enrichment of the cultures was undertaken to provide the most suitable growing condition for the algae, in order to have most of the strains growing well in a laboratory environment, were incorporated.

Only after these methods, the screening of the stains was done by isolation and identification of algal species capable of substantial lipid production, targeting organisms with rapid growth rate and tolerance to environmental parameters. The conventional method used for lipid determination involves solvent extraction and gravimetric determination [91]. Lipid analyses was only conducted after the strains were deemed suitable for high growth rate under extreme temperature conditions.

The authors then proceeded to identify the strains using a molecular approach of PCR amplifications of ITS regions of rRNA for genetic identification [78]. From this study the authors widened their focus from the production of biofuels from microalgae at the downstream level of extracting oil, but to search for new and distinctive microalgae from a multitude of aquatic environmental conditions, and the potential of these strains to generate other products such as polyunsaturated fatty acids [92].

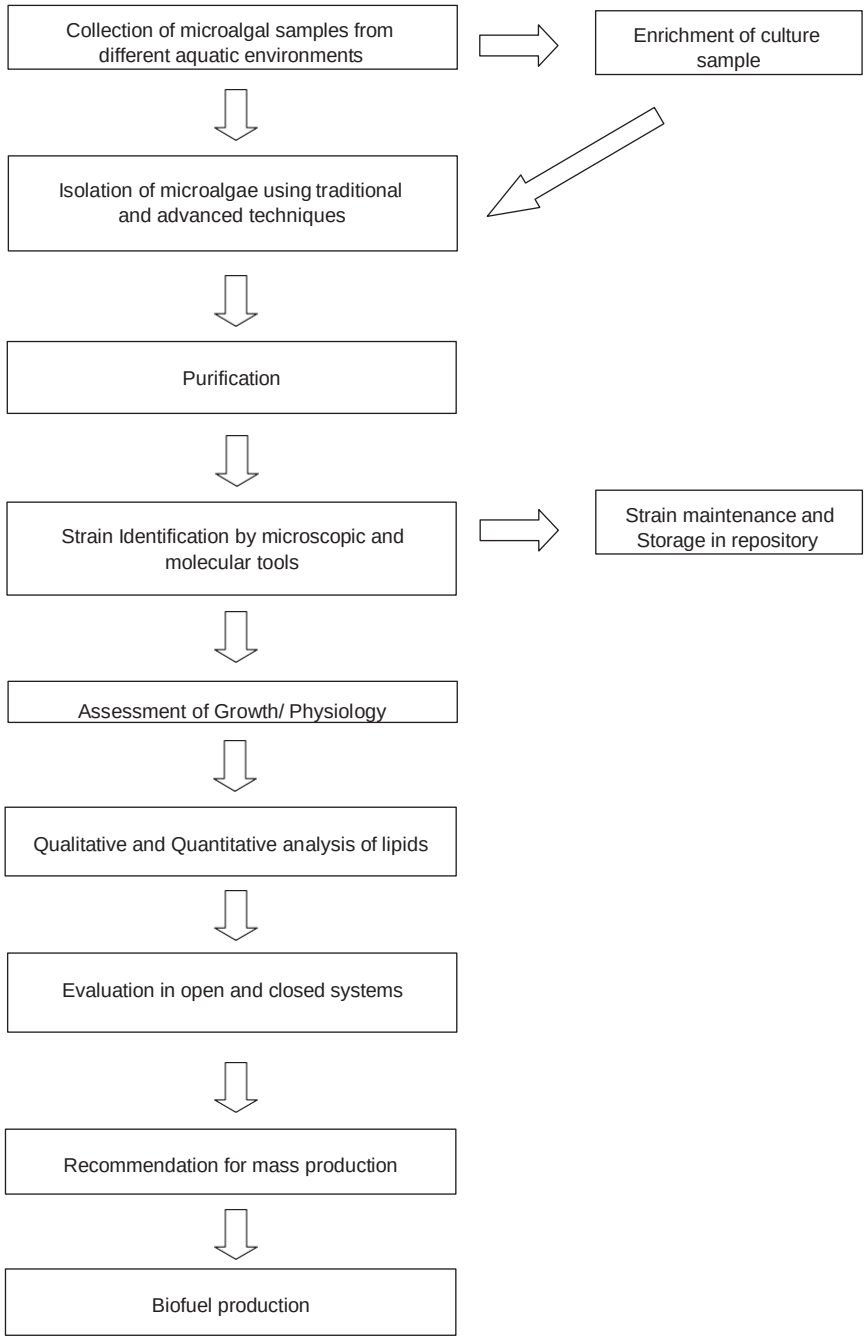


Figure 1.8: Steps involved and outcome of bioprospecting of microalgae for biofuels production [78]

1.8.3 Yeast for alcohol production

While this technique was suitable for algae, a large body of literature has also been invested in bioprospecting yeasts, though this has mainly focused on ethanol production to date. For example, in a typical study Lee et al. (2011) investigated the wild yeasts sampled from grapes from around the world and also from variety of fruits grown in South Korea, all bought from shops from Incheon and Seoul. They commenced with sampling of yeasts by isolating yeasts from the surface of fruits mentioned and suspended into yeast-peptone-dextrose (YPD) broth at pH 4.5. Lysozyme and the chloramphenicol were used at different point of time to eliminate any filamentous fungi and bacteria [93].

The aim of the screening was to isolate strains that are highly tolerant towards high glucose concentrations of 30 %, 40 %, and even 50 % at 30 °C, while still being able to produce high ethanol titres. Another parameter for selection was to determine their tolerance to high alcohol levels, attempted by culturing the isolates in YPD agar plates containing 15 % or 30 % alcohol. The next screening was for strains that possessed the ability to grow in 2% Maltose at 30 °C. The growing stains were than screened for α-amylase production. An alcohol analysis was later carried out for all surviving strains. These strains were than identified based on 26S rDNA sequencing, where the homologous sequences were matched in the NCBI database using the BLAST (Basic Local Alignment Search Tool, an algorithm for comparing primary biological sequence information, such as the nucleotides of DNA sequences) platform.

In another study by Choudhary, Singh, and Nai (2007) 637 isolates of wild yeasts were obtained from the earlier stage of bioprospecting from which 120 strains with at least one of the desired characteristics were identified. For example, yeasts with superior glucose and alcohol tolerance, demonstrated that they can endure high osmotic pressure and higher alcoholic content. A further trait was examined for their ability to metabolise maltose, to increase the amount of available sugar in a typical fermentation for the yeast. Selecting yeasts that could break down starch (with the availability of α-amylase), is another trait in developing fortified wine with higher alcohol content and smoother taste.

From all of these factors, *Pichia anomala* (later renamed *Wickerhamomyces anomalus*) and one strain of *Saccharomyces cerevisiae* were shown to be the highest performing strains and tolerant to these stresses [94]. Largely, the adeptness to effectively utilize maltose correlated with tolerance to osmotic pressure but not to alcohol tolerance.

The study was extremely successful in identifying yeast suitable for first generation processing from simple sugars, however, this did not produce yeasts that were suitable for growing on further lignocellulosic feedstocks. In an attempt to produce suitable ethanol producers from lignocellulosic hydrolysates *Choudhary et. al.* used a different methodology [94] .

In this research, yeasts were taken from rotten fruit samples and distillery waste samples in summer (April to June of 2014) where the temperature was between 35°C and 47°C. The collection was cultured in malt extract glucose yeast extract peptone (MGYP) medium, added with chloramphenicol. They were then identified using ITS region amplification and sequencing, then identification using the NCBI database. As a result, 10 strains were identified.

They were subjected to grow in the broth in 30°C and 40°C, afterwards. The most significant step by Choudhary et. al was to culture in different sugar contents at 0.5% w/v: xylose, L-arabinose, D-arabinose, galactose, cellobiose, mannose and rhamnose. All cultures were grown at 40 °C with oxygen available for 3-5 days. Inhibitors (furfural and 5-hydroxymethyl furfural) were also added to the broth. Yeasts were also subjected to an array of temperatures: 30°C, 35°C, 40°C, 42°C, 45°C. A summary of the results for the 40°C study is given in table 1.4.

Table 1.4: Sugar utilization patterns of various yeast strains at 40 °C after 3 days of incubation (O.D.) (taken from Choudhary et al. [94]).

Strain no.	Culture	d-Arabinose	l-Arabinose	Cellobiose	Galactose	Glucose	Mannose	Rhamnose	Xylose
1	*S. cerevisiae LN*	0	0	0	1.83 ± 0.03	1.85 ± 0.02	1.88 ± 0.02	0	0.08 ± 0.0
2	*S. cerevisiae DC*	0	0	0	1.55 ± 0.00	1.56 ± 0.00	1.56 ± 0.01	0	0.07 ± 0.0
3	*C. tropicalis JRC1*	0	0	1.42 ± 0.00	2.17 ± 0.01	2.08 ± 0.03	2.11 ± 0.05	0	2.17 ± 0.0
4	*P. kudriavazevii JRC2*	0	0	0	0	1.86 ± 0.02	1.91 ± 0.0	0	0.045 ± 0.0
5	*C. tropicalis JRC3*	0	0	1.48 ± 0.01	1.93 ± 0.0	1.91 ± 0.01	1.96 ± 0.02	0	1.9 ± 0.0
6	*P. kudriavazevii JRC4*	0	0	0	1.62 ± 0.05	2.13 ± 0.04	2.14 ± 0.02	0	0.37 ± 0.0
7	*S. cerevisiae JRC5*	0	0	0	1.79 ± 0.02	1.63 ± 0.04	1.57 ± 0.03	0	0
8	*S. cerevisiae JRC6*	0	0	0.24 ± 0.00	1.7 ± 0.02	1.56 ± 0.00	1.55 ± 0.04	0	0.18 ± 0.01
9	*C. tropicalis Y6*	0	0	1.9 ± 0.02	2.02 ± 0.01	1.98 ± 0.09	2.06 ± 0.07	2.04 ± 0.04	2.17 ± 0.05
10	*W. anomalus JRC7*	0	0	0	1.9 ± 0.06	2.0 ± 0.01	1.9 ± 0.06	0	0

All the yeast strains showed a slight longer lag time at 40°C. The results of the different sugar content of can be examined in table 1.2, significantly all 10 strains could not grow on either isomer of arabinose, and except for *C. tropicalis* Y6, all strains could not survive when rhamnose is the sole carbon source. The main outcomes from this is the identification of *Saccharomyces cerevisiae* JRC6, achieved highest fermentation efficiency (87.9%) at 40°C. This is an important characteristic for ethanol production from SSF lignocellulosic hydrolysates.

1.8.4 Screening for oleaginous yeasts.

While both of these studies were well designed, they focussed solely on ethanol production. Duarte et al. used bioprospecting to find suitable oleaginous yeasts from Brazil in 2013 [95]. Yeasts were collected from different Brazilian regions from soil, stems, fruits, and flowers. They were stored in MGYP at 5°C and then reactivated at 30°C. A simple screening for oil was carried out using Sudan Black B staining. Later the strains were grown in medium with 30 g/L of pure and raw glycerol as its sole carbon source. Later the strains were subjected to genetic identification of the D1/D2 domains subunit 26S of rDNA. Next, the biomass of the cultures was harvested and lipids were extracted using the conventional methods of Bligh and Dyer [91], after which, lipid were converted into FAME and analysed with FID detector.

Out of 129 strains, 42 showed existence of oil from the Sudan Black test. Although not precise, this technique proved to be a useful first screening method filter with only the worthwhile yeasts selected. Then, only five strains were selected and characterized as the better strains in the collection from qualitative analysis of Sudan Black Test.

In raw glycerol, the biomass was higher in weight than pure glycerol. Impurities in glycerol proven to be an advantage for cell growth and accumulation of lipids. The lipid content for one of the strains, *Candida sp*. LEB-M3 cultured was profiled. In raw glycerol, C18:2 (Linoleic acid, a polyunsaturated fatty acid) showed dominance while in pure glycerol, C18:1 (vaccenic acid, a monounsaturated) constituted the highest percentage. This study demonstrated the utilization of raw glycerol as carbon source in lipid accumulation in yeasts [95].

Another similar study investigated suitable yeasts for the production of lipid from lignocellulosic hydrolysates. They decided that inhibitors, produced from the breakdown of sugars would be the largest challenge [96]. To this end a range of yeasts were not collected from the wild but purchased from the General Microbiological Culture Collection Centre in Beijing, China and screened for activity. The fermentation medium used was a nitrogen-limited media with additions of other substrates as per needed.

In the first screening, glucose and xylose were used as the carbon sources, and added to the nitrogen-limiting medium. Only six strains survived this process and lipid extraction was only done on these strains (table 1.5).

Table 1.5: The dry cell mass (DCM), lipid in medium, lipid in cells and lipid yield of the six oleaginous yeast strains after cultured 96 h in the nitrogen-limited medium [96].

Microorganisms	DCM (g/L)	Lipid in medium (g/L)	Lipid in cells (wt %)	Lipid productivity (g/L h^{-1})	Lipid yield (g/100 g glucose)
T.cutaneum 2.1374	2.75	1.09	39.8	0.011	10.1
L. starkeyi 2.1608	9.35	2.04	21.8	0.097	10.5
L. starkeyi 2.1390	6.16	2.29	37.2	0.021	10.6
R. glutinis 2.107	4.01	0.52	13.0	0.042	4.92
R. glutinis 2.704	5.49	0.92	16.7	0.057	2.78
R. toruloides 2.1389	4.26	1.67	39.3	0.044	12.6

Table 1.5 shows the lipid as a percentage of the cell from the species, this ranged from 13% to 39.8%. The fatty acid composition profiled majorly palmitic acid (C16:0), stearic acid (C18:0), oleinic acid (C18:1), linoleic acid (C18:2) and palmitoleic acid (C16:1) [96].

Under the presence of inhibitors, the yeasts are put under high stress levels that they react differently to producing lipids. With organic acids, namely acetic acid, formic acid, and levulenic acid, the six strains screened earlier were cultured. Acetic acid and formic acid strongly inhibited all six selected strains. Levulinic acid, on the other hand, did not affect the strains even at high concentrations. Furfural had the greatest impact on all six strains, over 5-(hydroxymethyl)furfural (5-HMF). *Trichosporon cutaneum* 2.1374 demonstrated the highest tolerance to both furfural and 5-hydroxymethyl furfural (5-

HMF), though both *Lipomyces starkeyi* strains did not survive being cultured in 5-HMF [96].

Under the phenol derivatives originating from lignin, vanilin was shown to be the most potent, stunting the growth of all strains except *T. cutaneum* 2.1374. For Hydroxybenzaldehyde, the effect was not as strong as vanillin, but growth declined as concentration increased. Again *T. cutaneum* 2.1374 exhibited the most resilience [96].

Lastly, the strains were exposed to Corn Stover hydrolysates at high concentrations of 80%, 60%, 40% and 20%. Again, *T. cutaneum* 2.1374 grew the best, but failed to grow completely at 80% of hydroysates. This experiment demonstrated that most oleaginous strains are not suitable for biotechnological production, as they cannot produce lipid under inhibitory conditions [96].

1.8.5 Genomics Bioprospecting

Gene Prospecting is another type of Bioprospecting, and will undoubtedly have a substantial future in determining suitable organisms [97]. In gene prospecting, enzymes with new specific phenotypes are able to be have high throughput with the coupling of Biotechnology at minimum cost [79].

In the last 20 years, however, the isolation of novel compounds from traditional bioprospecting has plummeted considerably, mostly because of the repeated isolation of the already known compounds [98]. This suggests that we have expended the possibilities of flora and fauna as sources for new therapeutics. However, with the advancement in genomic studies, where genes and genomic sequencing have now becoming routines, this phenomena has been transformed [99].

Due to the high number of genes in some bacteria, for instance, not all proteins with high potential could be distinguished due to their low production. In a study of *Streptomyces coelicolor* A3 showed that genes encoding biosynthesis of compounds for example that express antibiotics were expressed at a very low level or none at all in laboratory settings, making the detection very problematic [100]. It was said that a unique combination of environmental factors possibly needed in the expression of a biosynthetic gene, making it not coded in certain conditions [101]. In studies, following

the earlier one, the genome sequences of two actinomycetes, *Streptomyces coelicolor* and *Streptomyces avermitilis*, uncover several series of evidently hidden natural-product gene clusters, suggesting that these well-learnt strains may produce a higher quantity of bioactive compounds than has been detected by fermentation culture analysis [102].

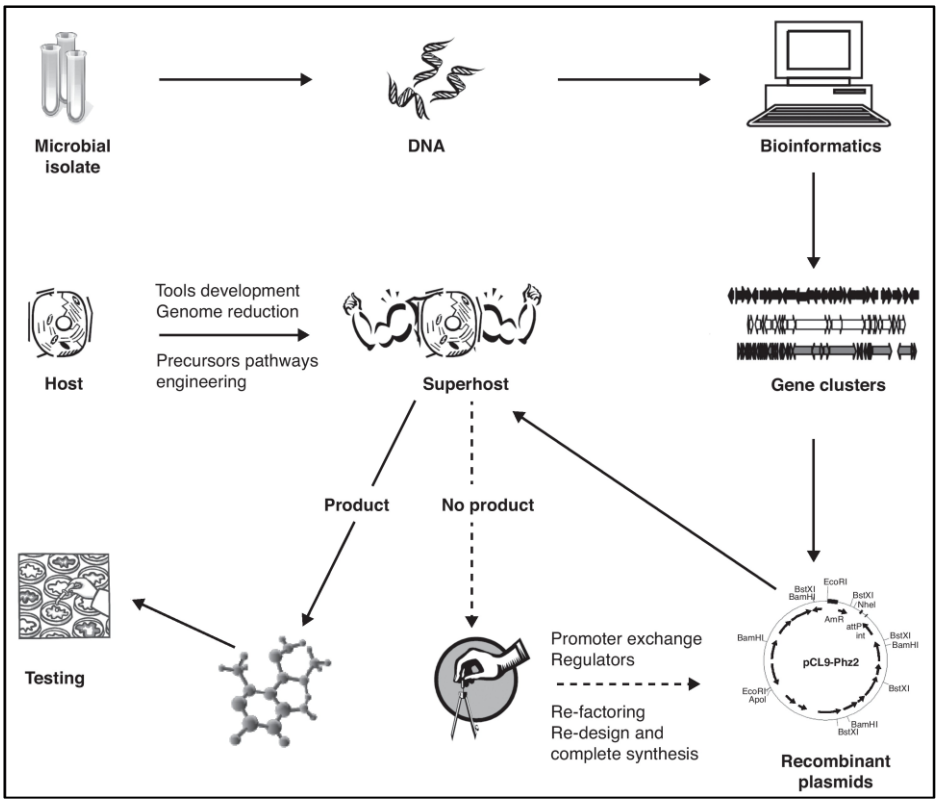


Figure 1.9: Overall tentative scheme for genome-based bioprospecting [99]

Zazopoulos et al. have established that standard fermentation broth analysis was unsuccessful to identify many strains that can produce enediyne antibiotics. The competency of actinomycetes to generate antibiotics and other bioactive chemicals has evidently been greatly underestimated. Their project established the worth of genome analysis in discovering cryptic metabolic products and directing coherent methods for the expression, detection, and purification of new bioactive natural products [103].

To have such studies possible, new bioinformatic tools is crucial as it allows well the detection of gene clusters of secondary metabolites in bacterial and fungal genomics, where the data is substantially large [104]. The mining of continually escalating number of genomic data, has led to the bioassay-independent discovery of gene clusters with

many important applications [105]. This information are then applied to the formation of recombinant plasmids containing that can be inserted into a tractable host [106].

1.9 Summary of literature

Palm oil production has a devastating effect on the environment, leading to wide scale deforestation and carbon emissions. Microbial oils, especially from yeast could be used as a substitute, but a number of factors must be obtained to realise this.

1. The lipid profile of the yeast must match that of palm oil, with elevated saturates present and the majority of the rest of the lipid being monounsaturated esters
2. The yeast must be able to be cultured on lignocellulosic residues, which means can use a wide array of sugars and be tolerant to inhibitors.

There is a wealth of natural biota that could be used for this effect, but to date there is no single bioprospecting methodology that has been applied successfully to wild yeasts to find suitable candidates for lipid production.

1.10 Aims and Objectives

The aim of this thesis is to develop a suitable bioprospecting methodology for the screening of yeasts for industrial lipid production, and select strains that can produce a palm oil substitute under inhibitory conditions on a range of sugars.

To achieve this aim, three specific objectives have been addressed each one presented in a separate chapter

The first objective is to develop a suitable bioprospecting technique, using a multiple gate plan to screen hundreds of strains from the local area. This will be designed to selected for pH tolerance, multiple sugar metabolism and inhibitor tolerance.

The second objective is to determine whether the selected strains can grow effectively under suitable conditions, with low pH and high inhibitors. On optimisation of the best conditions the yeasts will be screened for lipid production and assessed further for their lipid profile.

The final objective is to decipher the genomic sequence of one strain of the yeast to have a full database of its genome.

Chapter 2

Bioprospecting for suitable oleaginous yeast

2.1 Introduction

Bioprospecting is a systematic technique used to search for organisms, compounds, genes, whole designs or structures found in natural biota with an aim for potential product development. This is undertaken by biological observation or biophysical, biochemical, and genetic methods and should not disrupt the natural environment under study [78].

Thomas Eisner is largely recognized as developing a collaborative effort among conservationists, scientists, the pharmaceutical industry, and biodiversity-rich countries to develop products from biological diversity to aid in conserving the chemical wealth for future generations and also to generate income for its conservation [82]. Traditionally, chemical bioprospecting has been focused in screening for new drugs in pharmaceutical use has broaden to search for other chemicals, e.g. Bioethanol, due to the increasing demands for novel renewable energy-harvesting technologies and to introduce sustainable energy [106].

In this chapter, bioprospecting will be used to find suitable yeasts for the production of a palm oil substitute, similar to cocoa butter substitutes [107], but from lignocellulosic hydrolysates. Microorganisms offer a potentially more sustainable pathway to edible oils due to their ability to be grown in mass over shorter periods of time, and can be produced in high titres consistently [76]. Some yeast produce microbial lipids (single cell oils) that have merited industrial interest in the past due to the simple fatty acid profile and lipid properties [108]. In addition, these yeasts accrue lipids as triglycerides and can accumulate oil to a maximum of 80% of their dry weight [69, 109].

Oleaginous yeasts wide distribution suggests that various methods of oil build up could have evolved in different groups of yeasts. The identification of new species, and strains within a species, that are efficient lipid producers could hold the key to developing a cost effective palm oil alternative [110]. The classification of non-saccharomyces yeast that can be cultured on lignocellulosic hydrolysates is the primary aim of the bioprospecting [111].

In most yeast bioprospecting, a one-pot step in the initial culturing of the samples, is frequently used, such as using pure glycerol as the single carbon source for screening yeast for lipid production [94]. Multiple culturing methods were also used for selection of yeast to differentiate their abilities of utilizing different sugars or withstand high temperature [112].

However, to be cultured on lignocellulosic hydrolysates then the yeasts must be able to convert multiple hexose and pentose sugars as well as be able to handle the inhibitors commonly produced from the depolymerisation [113]. These inhibitors namely, acetic acid, furfural, 5-HMF and phenolic components are toxic to most yeasts, as well as some evidence suggesting that they reduce lipid production [53],[114].

2.2 Aims and objectives

The aim of this chapter is to develop a suitable bioprospecting technique for an oleaginous yeast that can be cultured on lignocellulose hydrolysates. A staggered culturing step will be used to ensure that the yeasts can handle multiple sugar sources and inhibitor concentrations as well as extremes of pH. This will be formulated to only select the most resilient from the hundreds of yeast strains gathered from fruits and flowers in the local are

2.3 Results and Discussion

2.3.1 Testing methodology

In order to select a resilient yeast strain, a protocol was designed to simulate the harshest conditions found in the conversion of depolymerised lignocellulose. To screen yeasts, two distinct pH conditions were tested (pH 4.8 and pH 9), a minimal nutritional media using different sugar components (xylose, glucose, arabinose and cellobiose) and inhibitors such as furfural, 5-HMF and organic acids. To select for the hardiest species, most common inhibitors produced during lignocellulose depolymerisation were used.

While it is tempting to produce one type of media with all these components included, this would not give a complete picture of a yeasts suitability. For example, one yeast might survive, while only metabolising the glucose. Therefore, a different approach, with multiple media, was taken (figure 2.1). To begin with any biota was rinsed with Ringer's solution to collect the prospective yeasts (and microbes) from the surface of the fruit and flower samples. Ringer's solution is isotonic and therefore this is a commonly used technique to collect samples in the field. The Ringer's solution does not alter the cells and ensures all existing species on the surface of the samples survive [115].

The yeast screening was run at two pH conditions (pH 4.8 and pH 9). By using pHs that are non-optimal for most bacterial growth, there is less risk of invasion from other microbes. Generally most invasive bacteria are known to prefer neutral pHs for optimal growth [116]. The spores of *Bacillus subtilis* for example cannot survive in strong acidic or basic conditions [117]. This would mean the yeast does not need to be contained and the waste streams do not need to be sterilised, this could potentially lead to a more cost effective process. In addition, the acidic pH is representative of the harsh environment in which the yeast will be subjected to lignocellulose degradation processes as organic acids can be produced from the degradation of sugars such as glucose and xylose.

Generally, most yeasts prefer acidic conditions and it is expected that the acidic conditions are a more likely path to a suitable biotechnological organism [118]. For example *S. cerevisiae* is commonly grown at pH 4.6 or below, Candida species tend to grow at below pH 5 [119], oleaginous yeasts such as *Yarrowia lipolytica* in pH 6 [120].

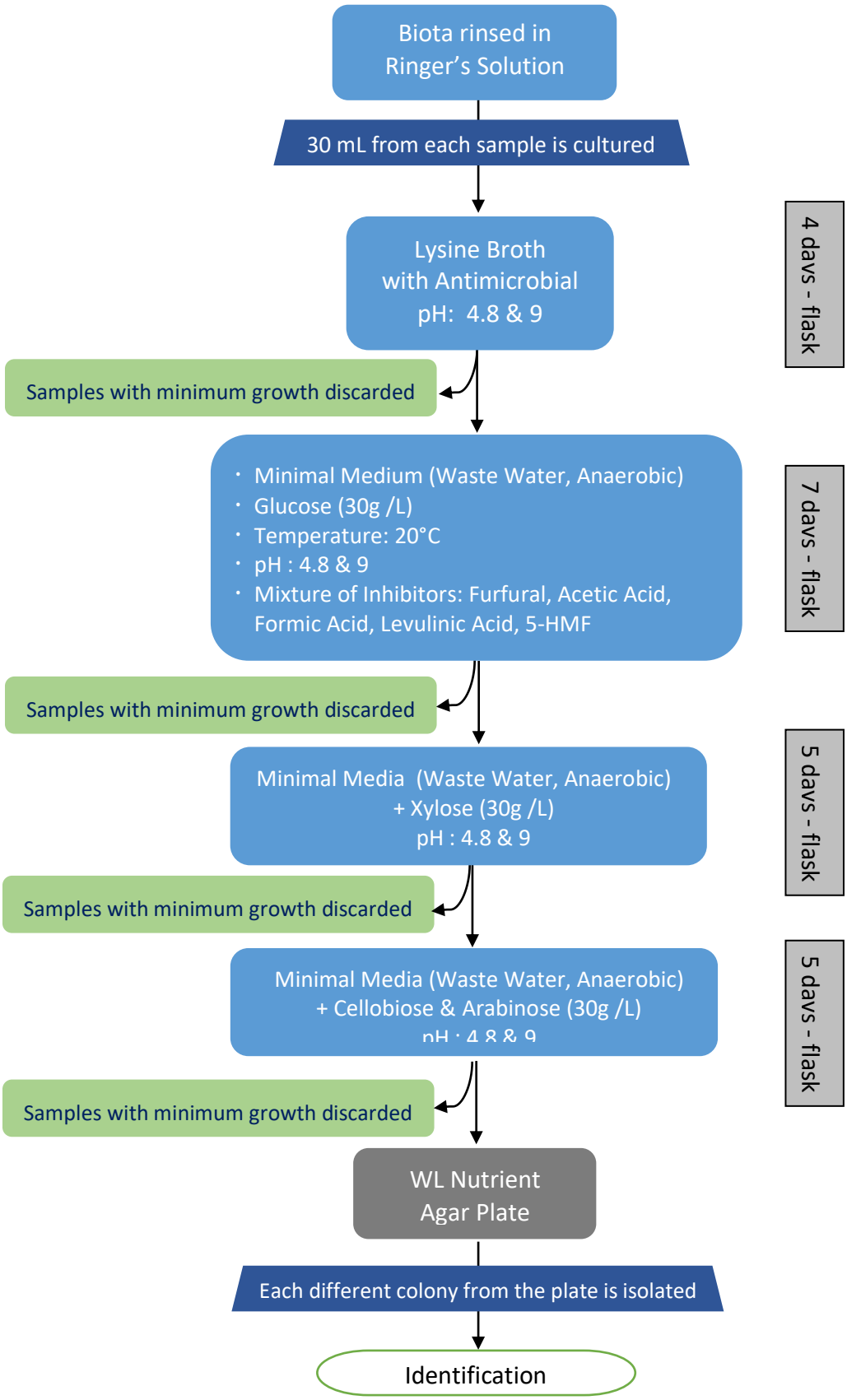


Figure 2.1 Experimental design flowchart: Laboratory stages to screen for the novel yeast.

In the first stage, minimal media with 30g/L of glucose and lysine (MML) was made at t2 pH's (the acidic media is given as MML_A, the basic as MML_B). The purpose of using lysine is to kill any *Saccharomyces sp.* that are present. Lysine is a growth inhibitor to almost all Saccharomyces yeast and various fungus [121], when used in nitrogen limited media. In *S. cerevisiae* the first step in the degradation of lysine is catalysed by saccharopine dehydrogenase with the formation of saccharopine. Three enzymes are involved in this first step: L-lysine ¢- aminotransferase, c-lysine s-dehydrogenase or acetyl CoA:L-lysine N-acetyl transferase, this varies in different species of organism [122]. *S. cerevisiae* may use more than one route to catabolise lysine. For example, when lysine is the sole source of nitrogen, the first degradation step is catalysed by an acetyl-transferase [121].

Under N-limiting growth conditions, L-lysine ¢-aminotransferase enzyme synthesis is blocked. Therefore, with no additional nitrogen present, *S. cerevisiae* cannot grow. Interestingly, this enzyme will also be repressed when excess of accumulative niitrogen is present [122]. Therefore, in the medium preparation at this initial stage, no NH_4, as an N-source, was included.

The second reason for the first stage was to exclude all bacteria. Two types of antibacterial chemical were used: Tetracycline and Ampicillin. Tetracycline is a wide spectrum antibiotic which targets both gram-positive and gram-negative bacteria. It is known to inhibit protein synthesis by blocking the attachment of aminoacyl-tRNA to the ribosomal acceptor (A) site [123]. Ampicillin, a semisynthetic group of *ß-lactams* antibiotics, is also a wide spectrum type antibiotic which inhibits peptidoglycan, a continuous covalent macromolecular structure found on the outside of the cytoplasmic membrane, synthesis on the cell wall [124]. Ampicillin is used in synergy with other antimicrobials to enhance the bacterial control [125].

After the first stage no attempt was made to identify the organisms, rather everything that survived over the timeframe was transferred to the second stage for cultivation. This was achieved by using 1 ml from this culture as the inoculum for the next.

The second stage in the screening was to test the ability of the yeasts to survive a harsh inhibitory environment that simulate the conditions of the lignocellulose degradation process the yeast will encounter (MMI). In this case the minimal media was used with 30 g l^{-1} glucose and a range of inhibitors namely: acetic acid, formic acid, levullinic acid, 5-hydroxymethyl furfural (5-HMF) and furfural. These chemicals are produced by the degradation of hexose and pentose sugars, creating an extremely harsh environment for any potential organism [107]. In MMI the nitrogen source was changed to NH_4^+.

The inhibitor concentration was chosen to be high, though realistic, with values that have been shown to be detrimental to most microbial species [108]. While some yeasts, such as *M. pulcherrima* have been shown to have a high resistance to these inhibitors separately by combining the inhibitors together then any synergistic effects would also be accounted for. Therefore, it was reasoned that any yeast that can survive the high levels of each inhibitor in addition to any synergistic effects would need to be a very resilient species.

Any yeast surviving stage two would then be carried on to stage 3. Other than inhibitors, the depolymerised lignocellulose comprises of many alternative different sugars. Cellulose and hemicellulose are made up of hexose and pentose based polymers. While celluloses are polymers of glucose, hemicelluloses are more heterogeneous and include pentoses (β-D-xylose, α-L-arabinose) and the hexoses; β -D-mannose, β -D-glucose, α -D-galactose. Though trace amounts of other sugars such as α-L-rhamnose and α -L-fructose are also observed [61].

Aside from glucose the major sugars produced from the depolymerisation of lignocellulose are xylose, cellobiose & arabinose [109]. An optimal organism for biotechnology would be capable of metabolising all the sugars present from depolymerised cellulose. To this end, two more stages were added to the screening program. In the first any yeast that survived the inhibitor stage was grown in minimal media with only xylose as the sole carbon source (MMX). Organisms that were capable of this were then subjected to a minimal medium containing arabinose & cellobiose as the sole carbon-source (MMAC).

At the end of the screening procedure, yeast colonies that demonstrated suitable growth were streaked onto a Wallerstein Laboratory (WL) Nutrient Agar. Different colonies growth on the WL agar will show distinctive colour or texture to easily distinguish one strain from another. These strains were isolated and grown on separate YPD agar plates.

Finally, each successful growth was identified using two different method of yeast species identification. The first procedure is using a yeast identification kit, API® ID 32 strip from bioMérieux, this was then replaced with a PCR Sequencing Method.

2.3.2 Testing the experimental design using *M. pulcherrima*

Previous work has demonstrated that *M. pulcherrima* is a promising yeast for biotechnology and is resistant to the severe environments produced by the depolymerisation of lignocellulose [129]. For example, it has been shown that *M. pulcherrima* can grow on wheat straw hydrolysates effectively and has good resistance to individual inhibitors [130]. Therefore, the effectiveness of the screening method was tested on *M. pulcherrima* first to determine the suitability of the screening method. Each stage of the program was carried out for up to 9 days. Though this was seen as excessive it gave a good indication of the optimal time that each stage should be run for when testing the unknown yeasts.

M. pulcherrima grew positively in the media. Normal growth curves were exhibited by the yeast in both acidic and basic state (fig. 2.2a). This is somewhat surprising as most yeasts have a narrow optimal pH range, and *M. pulcherrima* has been shown previously to thrive at low pH [119]. However, while the pH of the MML_A remained fairly constant between pH 4 and pH 5, the basic MML_B is reduced from 9 to a neutral pH. The main reason for culturing at an extreme pH is to reduce the likelihood of a bacterial invasion, where if the pH is reduced to neutral it becomes counterproductive.

To attempt to keep the pH in the alkaline region, a buffer is needed. In this case TRIZMA® is a strong basic buffer to counter the changes and was used in this investigation.

No bacterial or *Saccharomyces cerevisiae* contamination was observed, which suggested that the combination of lysine and antibiotics is suitable for culturing these

yeasts.hough it should be noted that *M. pulcherrima* also produces a range of antimicrobials such as pulcherrimin that could also be aiding in reducing any contamination [72].

The growth in the basic media was better than in the acidic media as shown in Figure 2.2. A higher rate of growth of the yeast in the basic media can be attributed to the fact that the pH is brought down to nearly pH 7. *M. pulcherrima* grows best at between pH 5.0 and 7.5 [131]. However, pH 7 would promote bacterial growth when there are no antimicrobials present, where pH's between 3.5-5 is more suitable in an industrial setting than the basic conditions [132].

Both of the cultures of *M. pulcherrima* in acidic and basic conditions had an extended lag time, of 48 hours, before entering the exponential phase. This extended lag time would be costly if run on an industrial process and a fast growing species is essential in limiting the expenditure to producing biological products. One reason for this might be the sole N source being lysine. This is potentially due to the yeast needing to break lysine down into ammonium, whereas in an industrial process there would be a suitable level of ammonium to begin with. When ammonium is present, under optimal conditions, *M. pulcherrima* has been shown to reach stationary phase in less than 72 hours [132]. A very good growth rate is shown by *M. pulcherrima* in this particular media producing roughly 3 g/L dry weight.

M. pulcherrima was then cultured in minimal media containing glucose and the inhibitors. Surprisingly, little growth was observed for *M. pulcherrima* on MMI_A (fig. 2.2b), despite previous work demonstrating that *M. pulcherrima* was resistant against all types of inhibitors at these concentrations and pH. This suggests that there is an adverse synergetic effect when all the inhibitors are present.

In a previous study, individual inhibitors were tested rather than mixing the inhibitors from different origin in one batch, and this demonstrated a far lower detrimental effect [133]. In many cases inhibitors have been observed to decrease ethanol yield with the phenolic compounds being the most inhibiting [133].

A synergistic effect was observed by Oliva et al [134]. In the culturing of *Kluyveromyces marxianus*, a thermotolerant yeast, there was no harmful effect on growth and production of ethanol when inhibitors were added as single compounds to the culture. However, the combination of different inhibitors: acetic acid (0–10 g/L); furfural (0–2 g/L); and catechol (0–1 g/L), substantially affected the growth and fermentation of *K. marxianus*. This demonstrated that the combination has a synergistic effect with respect to ethanol yield and culture growth as opposed when grown in the compounds separately [135].

However, when grown in MMI_B *M. pulcherrima* hits the exponential phase after 120 hours (as shown in Figure 2.2b), even in this harshest condition, *M. pulcherrima* still survived. The pH also remains relatively constant not going down to a near neutral pH maintaining an axenic condition.

To test the effectiveness of the growth of xylose, *M. pulcherrima* was cultured in minimal medium containing only xylose as the sugar source (fig. 2.2c). Under the acidic conditions, growth was observed though *M. pulcherrima* showed a slower growth rate compared to culturing using glucose. However, when grown in basic conditions, almost zero growth was observed. This was to be expected, as previous studies have shown reduced growth rates on xylose [72]. This was not seen as a major issue here, as the yeasts do not need to thrive under all conditions rather demonstrate an ability to catabolise the sugars.

Saccharomyces cerevisiae does not ferment D-xylose naturally, though a number of alternative yeasts such as *Pichia stipitis, Candida shehatae* and *Pachysolen tannophilus* all can [60]. Santomauro et al., demonstrated that *M. pulcherrima* can grow on xylose, though the growth rate was far slower than with glucose. It needs a longer adaptation time to produce the relevant enzymes though the adaption is reasonably short if there are high levels of glucose also present.

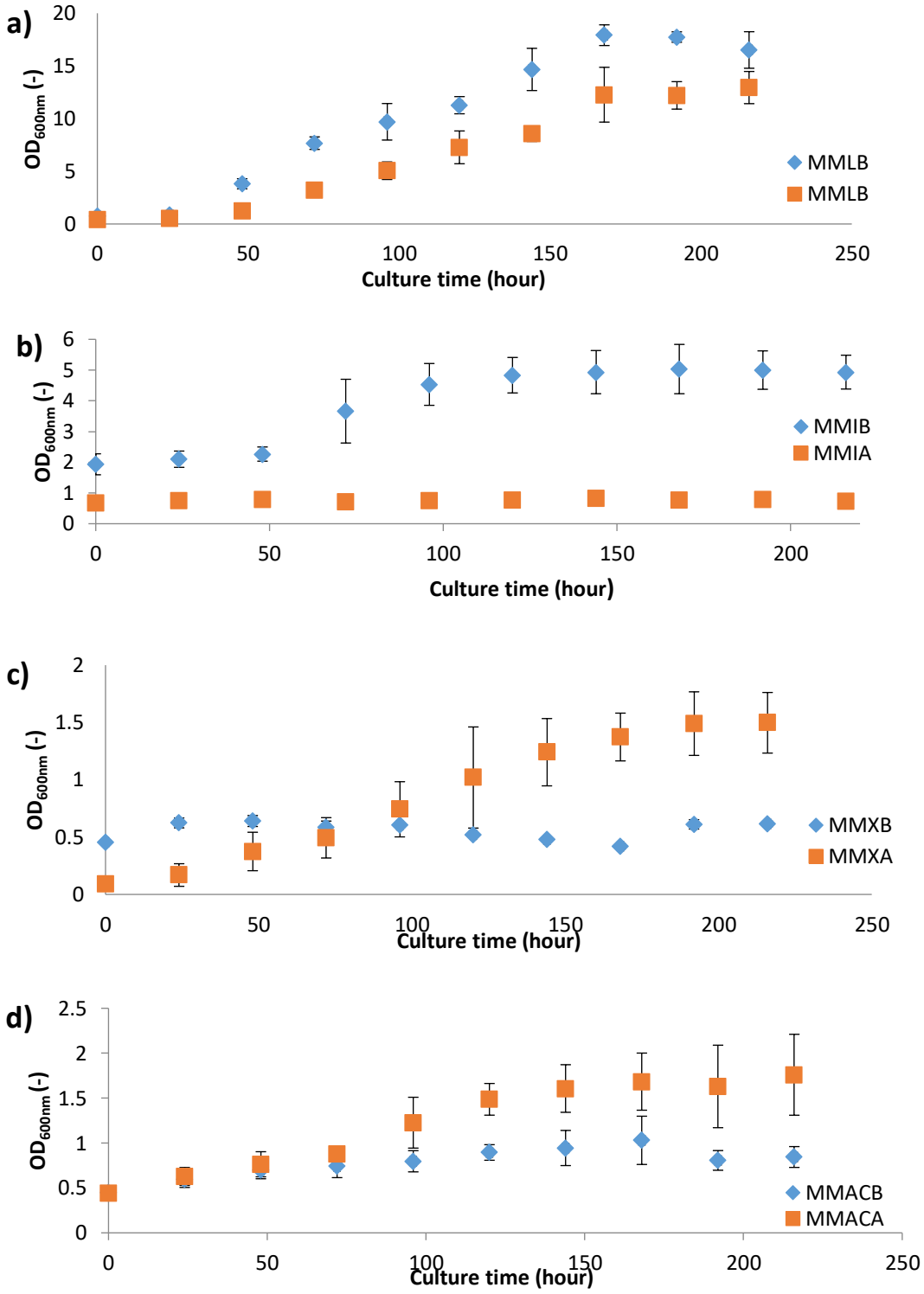


Figure 2.2: a) Growth of *M. pulcherrima* in an acidic minimal medium (MML_A) and basic minimal media (MML_B) containing glucose, lysine & antibiotics. b) Growth of *M. pulcherrima* in both an acidic (MMI_A) and basic (MMI_B) minimal medium with glucose and inhibitors (furfural (10mM), acetic acid (60mM), formic acid (60mM), levullinic acid (60mM), 5-HMF (10mM)). c) Growth of *M. pulcherrima* in both an acidic (MMX_A) and basic (MMX_B) minimal medium with xylose (30 g/L). d) Growth of *M. pulcherrima* in minimal medium with arabinose, 15 g/L and cellobiose, 15 g/L, as the sole carbon source.

In the last stage, *M. pulcherrima* was cultured in minimal medium however the carbon source was a combination of 15 g/L arabinose and 15 g/L cellobiose. Similarly, to the growth observed with MMX under acidic conditions the growth rate was slow when cultured with arabinose and cellobiose (fig. 2.2d), while there was only 0.1 difference of growth in O.D. reading under the basic conditions. The pH for both acid & base changed slightly from there starting points, indicative of low metabolic activity. Similarly to catabolising xylose, to use cellobiose (an isomer of sucrose) as a carbon source, the enzyme cellubiase is needed [137]. It was proven at temperature higher than 40°C and pH lower than 7 they would be the most optimum [138]. In other species such as *Candida albicans*, D-arabinose is used as a carbon source by firstly reducing it to D-arabitol. D-arabitol is then converted to D-ribulose by NAD- dependent D-arabitol dehydrogenase (ArDH). The last step, which is still postulated, D- ribulose could be phosphorylated at C-5 by a ribulokinase and further metabolized in the pentose phosphate pathway [139, 140]. As *M. pulcherrima* can grow on both pentoses, it seems likely that it also can produce these enzymes.

M. pulcherrima grew reasonably well in all of the screening cultures except in the inhibitors (MMI). It seems that while the standard lab strain of *M. pulcherrima* would not have passed the tests, a hardier yeast than *M. pulcherrima* would as we will prove in the following experiments. Based on these results it was reasoned that these protocols were suitable for screening yeasts, though would not give more survivors than could effectively be investigated.

2.3.3 Bioprospecting for Novel Yeasts

On the successful development of a suitable bioprospecting screening protocol, a local vineyard was chosen as the first screening site. For this, Mumford's Vineyard, located only 5.6 miles away from the University was chosen as the sampling ground due to the close proximity (fig. 2.3). At the time of sampling the grapes were approaching harvesting age. At this point, the grapes have ripen and acidic juice would seeped onto the fruit surface and provide sugar for the yeast's growth. A second set of sample collection was done on the campus of the University of Bath itself. A small sample was also collected in the outskirts of Bath (Lansdown).

Figure 2.3: Satellite photograph of Mumford's Vineyard, with sampling sites highlighted

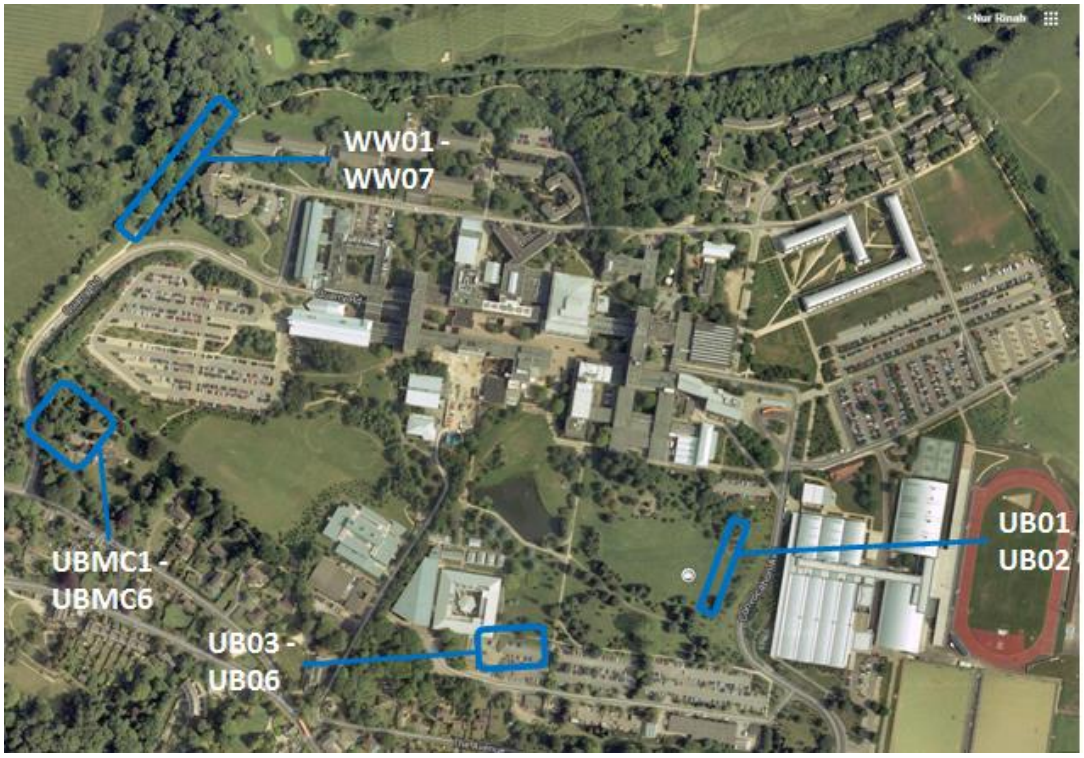


Figure 2.4: Aerial photograph of the University of Bath campus, with sampling sites highlighted

Figure 2.5: Aerial photograph of Lansdown Racecourse, with sampling sites highlighted

Table 2.1 : Description of where the samples were collected from in Muimford's Vineyard.

No	Sample Description	Section
1	Large red grapes	1(a)(b)
2	Small red grapes	1(a)(b)
3	Large red grapes	1(a)(b)
4	Large red grapes	1(a)(b)
5	Medium red grapes	1(a)(b)
6	Small red grapes	1(a)(b)
7	Medium red grapes	1(a)(b)
8	Small red grapes	1(a)(b)
9	Small green grapes	1(a)(b)
10	Small yellow flowers	1(a)(b)
11	Medium red grapes	1(a)(b)
12	Medium red grapes	1(a)(b)
13	Small red grapes	1(a)(b)
14	Small red grapes	2
15	Small green grapes	2
16	Small green grapes	2
17	Small yellow flowers	2
18	Small green grapes	2
19	Small red grapes	2
20	Small green grapes	2
21	Small red grapes	2
22	Medium red grapes	2
23	Small green grapes	2
24	Small green grapes	2
25	Small yellow flowers	2
26	Small yellow flowers	2

Table 2.2 : Description of where samples were collected in University of Bath & Bath area

No	Label	Sample Description	Location
1	UB01	Strawberry Apple	In front of Sports Training Centre
2	UB02	Firethorn (*Pyracantha coccinea)*	In front of Sports Training Centre
3	UB03	Yew (*Taxus baccata*)	3S Carpark
4	UB04	Yew (*Taxus baccata*)	3S Carpark
5	UB05	Ivy (Hedera helix)	3S Carpark
6	UB06	Ivy (Hedera helix)	3S Carpark
7	UBMC1	Blackberry (*Rubus fruticosus* agg.)	Vicinity of University Medical Centre
8	UBMC2	Blackberry (*Rubus fruticosus* agg.)	Vicinity of University Medical Centre
9	UBMC3	Blackberry (*Rubus fruticosus* agg.)	Vicinity of University Medical Centre
10	UBMC4	Yew (*Taxus baccata*)	Vicinity of University Medical Centre
11	UBMC5	Purple wild flower	Vicinity of University Medical Centre
12	UBMC6	Plum (*Prunus domestica*)	Vicinity of University Medical Centre
13	UBGC1	Blackberry (*Rubus fruticosus* agg.)	Roadside near Golf Club Carpark
14	UBGC2	Blackberry (*Rubus fruticosus* agg.)	Roadside near Golf Club Carpark
15	UBGC3	Elderberry (*Sambucus nigra)*	Roadside near Golf Club Carpark
16	UBGC4	Elderberry (*Sambucus nigra)*	Roadside near Golf Club Carpark
17	WW01	Blackberry (*Rubus fruticosus* agg.)	Behind Westwood Residence
18	WW02	Blackberry (*Rubus fruticosus* agg.)	Behind Westwood Residence
19	WW03	Blackberry (*Rubus fruticosus* agg.)	Behind Westwood Residence
20	WW04	Hawthorn (Crataegus monogyna)	Behind Westwood Residence
21	WW05	Hawthorn (Crataegus monogyna)	Behind Westwood Residence
22	WW06	Elderberry (Sambucus nigra)	Behind Westwood Residence
23	WW07	Blackberry (*Rubus fruticosus* agg.)	Behind Westwood Residence
24	LD01	Blackberry (*Rubus fruticosus* agg.)	Lansdowne Race Course
25	LD02	Elderberry (*Sambucus nigra)*	Lansdowne Race Course

2.3.4 Sampling and preparation of the yeast strains

Bioprospecting sampling was carried out at the Mumford's Vineyard, Bath. This vineyard is a 1.5 hectare plot overlooking the Avon valley east of Bath. In the vineyard various grapes are grown next to one another these include Kerner; Madeleine Angevine; Triomphe d'Alsace; Leon Millot and Reichensteiner varieties.

The grapes and flowers were randomly sampled, then, aseptically handpicked directly from the vineyard (vineyard sample), one week prior to commercial harvesting period. The grapes or flowers were carefully picked, so as not to disturb the biota.

A second sampling was done in March of 2014. For this sampling the vicinity of the University of Bath was used where wild berries and others grow in abundance. The ripening berries were aseptically plucked from the plants. Immediately, all samples are packed into sterilised plastic falcon tubes prior to delivery to the lab.

In the lab, samples were rinsed in Ringer's solution. And transferred directly to both acidic and basic MML broths in individual 250-ml Erlenmeyer flask. Each sample was then analysed daily for the optical density (O.D.) to assess the growth of the culture. Random samples were taken daily and analysed under the microscope to confirm that there was no bacterial growth.

2.3.5 Application of the protocol

2.3.5.1 First Collection

Stage 1, growth of unknown biota on MML

26 tubes of samples were collected aeseptically from the vineyard (fig 2.3, table 2.1). Each sample was rinsed carefully with Ringer's solution. The Ringer's solution with the captured microbiota was then transferred directly to the MML broth.

Each sample was then analysed daily for the optical density to assess the growth of the culture (Table 2.3 a & b). Almost all samples contained yeasts or moulds exhibited excellent growth in the acidic media. No bacteria or *Sacchromyces sp.* were observed for these cultures.

The yeast cultures also grew in the basic medium (MML_B), though less effectively than in the acidic medium (MML_A). This demonstrates that most yeasts prefer acidic

conditions. For both the culture conditions, with two comparatively minimum growth in samples 1B & 3B. The experiments were stopped after 4 days, as to be an effective biotechnological organism, the yeasts must grow quickly.

At the end of day 4, abundance of microbes were seen growing indicated by the darker yellow and green shades (table 2.3). Each of the samples had more than one surviving species. It should also be noted that at this point there is no way to differentiate between samples, therefore it is reasonable to assume that the same strain can be dominant in a number of the cultures.

Stage 2, growth of unknown biota on MMI

The survivors from stage 1 were then used to inoculate MMI, under both acidic and basic conditions. Due to the slow growth observed for *M. pulcherrima* the cultures were held for 7 days as opposed to 4 days, in case the lag time was substantial. In both the acidic and basic cultures the lag time was found to be at least 4 days for the yeasts that grew well. Yeasts and other microbes need time to adjust to the toxicity [110]. Again, more yeasts survived in acidic medium compared to the basic medium. This is expected, as grapes furnish an acidic environment to these acidophilic yeasts.

This step is far harsher than the previous stage and most of the yeasts could not grow under these conditions. In MMI_A, 10 samples out of the 26 taken were able to grow significantly under these conditions shown in shades of orange and green on table MMI. We chose the samples with O.D. reading greater then 3.0 as the indication of a healthy growth with the longer lag phase.

Under the basic conditions less of the collection survived the inhibitory stage. Samples 12B and 23B demonstrated considerable growth where their O.D.'s are the highest in this experiment.

As there are a number of acidic inhibitors, acidophilic yeast would mostly adapt easily and can survive these conditions, hence the higher number of survivors in MMI_A. The samples that survived in MMI_A and MMI_B are not all overlapping and presumably represent different species.

(a)	Location of sampling (MML_A)	OD_{600nm} (-)	
		Day 0	**Day 4**
Section 1(a)(b) Mumjford's Vineyard (1A)		0.257	8.16
Section 1(a)(b) Mumjford's Vineyard (2A)		0.257	8.26
Section 1(a)(b) Mumjford's Vineyard (3A)		0.257	8.28
Section 1(a)(b) Mumjford's Vineyard (4A)		0.257	18.86
Section 1(a)(b) Mumjford's Vineyard (5A)		0.257	5.13
Section 1(a)(b) Mumjford's Vineyard (6A)		0.257	6.93
Section 1(a)(b) Mumjford's Vineyard (7A)		0.257	14.06
Section 1(a)(b) Mumjford's Vineyard (8A)		0.257	3.42
Section 1(a)(b) Mumjford's Vineyard (9A)		0.257	7.45
Section 1(a)(b) Mumjford's Vineyard (10A)		0.257	11.50
Section 1(a)(b) Mumjford's Vineyard (11A)		0.257	5.80
Section 1(a)(b) Mumjford's Vineyard (12A)		0.257	5.63
Section 1(a)(b) Mumjford's Vineyard (13A)		0.257	15.90
Section 2 Mumford's Vineyard (14A)		0.257	19.72
Section 2 Mumjford's Vineyard (15A)		0.257	17.46
Section 2 Mumjford's Vineyard (16A)		0.257	28.28
Section 2 Mumjford's Vineyard (17A)		0.257	6.56
Section 2 Mumjford's Vineyard (18A)		0.257	8.58
Section 2 Mumjford's Vineyard (19A)		0.257	17.02
Section 2 Mumjford's Vineyard (20A)		0.257	4.68
Section 2 Mumjford's Vineyard (21A)		0.257	13.86
Section 2 Mumjford's Vineyard (22A)		0.257	12.24
Section 2 Mumjford's Vineyard (23A)		0.257	23.76
Section 2 Mumjford's Vineyard (24A)		0.257	14.10
Section 2 Mumjford's Vineyard (25A)		0.257	19.40
Section 2 Mumjford's Vineyard (26A)		0.257	14.44

(b)	Location of sampling (MML_B)	OD_{600nm} (-)	
		Day 0	Day 4
Section 1(a)(b) Mumjford's Vineyard (1B)		0.242	0.66
Section 1(a)(b) Mumjford's Vineyard (2B)		0.242	17.02
Section 1(a)(b) Mumjford's Vineyard (3B)		0.242	0.77
Section 1(a)(b) Mumjford's Vineyard (4B)		0.242	6.21
Section 1(a)(b) Mumjford's Vineyard (5B)		0.242	4.08
Section 1(a)(b) Mumjford's Vineyard (6B)		0.242	18.98
Section 1(a)(b) Mumjford's Vineyard (7B)		0.242	14.28
Section 1(a)(b) Mumjford's Vineyard (8B)		0.242	11.76
Section 1(a)(b) Mumjford's Vineyard (9A)		0.242	13.18
Section 1(a)(b) Mumjford's Vineyard (10B)		0.242	8.76
Section 1(a)(b) Mumjford's Vineyard (11B)		0.242	5.55
Section 1(a)(b) Mumjford's Vineyard (12B)		0.242	3.88
Section 1(a)(b) Mumjford's Vineyard (13B)		0.242	10.72
Section 2 Mumjford's Vineyard (14B)		0.242	17.90
Section 2 Mumjford's Vineyard (15B)		0.242	13.20
Section 2 Mumjford's Vineyard (16B)		0.242	7.96
Section 2 Mumjford's Vineyard (17B)		0.242	19.82
Section 2 Mumjford's Vineyard (18B)		0.242	1.064
Section 2 Mumjford's Vineyard (19B)		0.242	17.34
Section 2 Mumjford's Vineyard (20B)		0.242	17.34
Section 2 Mumjford's Vineyard (21B)		0.242	8.34
Section 2 Mumjford's Vineyard (22B)		0.242	13.02
Section 2 Mumjford's Vineyard (23B)		0.242	3.54
Section 2 Mumjford's Vineyard (24B)		0.242	12.12
Section 2 Mumjford's Vineyard (25B)		0.242	2.96
Section 2 Mumjford's Vineyard (26B)		0.242	8.56

Table 2.3: (a) Growth (OD_{600nm}) of the unidentified cultures 1-26 taken from the vineyard in MML_A, starting pH 4.8 at day 0 and endpoint (OD_{600nm}) at day 4.

(b) Growth (OD_{600nm}) of the unidentified cultures 1-26 taken from the vineyard in MML_B, starting pH 9 at day 0 and endpoint (OD_{600nm}) at day 4.

(a)	Location of sampling (MMI_A)	OD_{600nm} (-)	
		Day 0	Day 7
Section 1(a)(b) Mumjford's Vineyard (1B)		0.104	3.33
Section 1(a)(b) Mumjford's Vineyard (2B)		0.104	1.98
Section 1(a)(b) Mumjford's Vineyard (3B)		0.104	6.00
Section 1(a)(b) Mumjford's Vineyard (4B)		0.104	1.60
Section 1(a)(b) Mumjford's Vineyard (5B)		0.104	1.62
Section 1(a)(b) Mumjford's Vineyard (6B)		0.104	1.83
Section 1(a)(b) Mumjford's Vineyard (7B)		0.104	2.06
Section 1(a)(b) Mumjford's Vineyard (8B)		0.104	1.97
Section 1(a)(b) Mumjford's Vineyard (9A)		0.104	1.57
Section 1(a)(b) Mumjford's Vineyard (10B)		0.104	1.82
Section 1(a)(b) Mumjford's Vineyard (11B)		0.104	1.44
Section 1(a)(b) Mumjford's Vineyard (12B)		0.104	10.46
Section 1(a)(b) Mumjford's Vineyard (13B)		0.104	2.50
Section 2 Mumjford's Vineyard (14B)		0.104	2.40
Section 2 Mumjford's Vineyard (15B)		0.104	3.20
Section 2 Mumjford's Vineyard (16B)		0.104	2.90
Section 2 Mumjford's Vineyard (17B)		0.104	2.33
Section 2 Mumjford's Vineyard (18B)		0.104	3.78
Section 2 Mumjford's Vineyard (19B)		0.104	2.45
Section 2 Mumjford's Vineyard (20B)		0.104	2.48
Section 2 Mumjford's Vineyard (21B)		0.104	5.56
Section 2 Mumjford's Vineyard (22B)		0.104	2.80
Section 2 Mumjford's Vineyard (23B)		0.104	9.1
Section 2 Mumjford's Vineyard (24B)		0.104	2.66
Section 2 Mumjford's Vineyard (25B)		0.104	1.61
Section 2 Mumjford's Vineyard (26B)		0.104	2.22

(b)	Location of sampling (MMI_B)	OD_{600nm} (-)	
		Day 0	Day 7
Section 1(a)(b) Mumjford's Vineyard (1A)		0.036	0.60
Section 1(a)(b) Mumjford's Vineyard (2A)		0.036	17.44
Section 1(a)(b) Mumjford's Vineyard (3A)		0.036	8.26
Section 1(a)(b) Mumjford's Vineyard (4A)		0.036	5.40
Section 1(a)(b) Mumjford's Vineyard (5A)		0.036	0.37
Section 1(a)(b) Mumjford's Vineyard (6A)		0.036	15.66
Section 1(a)(b) Mumjford's Vineyard (7A)		0.036	0.73
Section 1(a)(b) Mumjford's Vineyard (8A)		0.036	0.30
Section 1(a)(b) Mumjford's Vineyard (9A)		0.036	15.00
Section 1(a)(b) Mumjford's Vineyard (10A)		0.036	0.71
Section 1(a)(b) Mumjford's Vineyard (11A)		0.036	1.06
Section 1(a)(b) Mumjford's Vineyard (12A)		0.036	0.33
Section 1(a)(b) Mumjford's Vineyard (13A)		0.036	5.84
Section 2 Mumford's Vineyard (14A)		0.036	0.46
Section 2 Mumjford's Vineyard (15A)		0.036	0.62
Section 2 Mumjford's Vineyard (16A)		0.036	0.60
Section 2 Mumjford's Vineyard (17A)		0.036	0.52
Section 2 Mumjford's Vineyard (18A)		0.036	0.51
Section 2 Mumjford's Vineyard (19A)		0.036	0.56
Section 2 Mumjford's Vineyard (20A)		0.036	0.45
Section 2 Mumjford's Vineyard (21A)		0.036	1.4
Section 2 Mumjford's Vineyard (22A)		0.036	9.69
Section 2 Mumjford's Vineyard (23A)		0.036	0.46
Section 2 Mumjford's Vineyard (24A)		0.036	0.70
Section 2 Mumjford's Vineyard (25A)		0.036	0.87
Section 2 Mumjford's Vineyard (26A)		0.036	6.28

Table 2.4: (a) Growth (OD_{600nm}) of the unidentified cultures 1-26 taken from the vineyard in MMI_A, starting pH 4.8 at day 0 and endpoint (OD_{600nm}) at day 7.

(b) Growth of the cultures in MMI_B which had a starting pH of 9 at day 0 and endpoint (OD_{600nm}) at day 7.

Stage 3, growth of unknown biota on MMX

From the previous medium containing Inhibitors, only the survivors were transferred to this minimal media containing only xylose as its carbon source (MMX). Surprisingly, while none of the yeasts grew to the same extent in xylose than they did in glucose, most of the strains showed some ability to grow on MMX in both basic and acidic conditions (fig. MMX). It seems probable that these yeasts would be able to assimilate xylose effectively in lignocellulose hydrolysate.

Under acidic conditions, almost all samples showed some growth as shown in figure 2.6 a, with only two samples in particular showing very little increase (6A and 22A). Sample 3A on the other hand demonstrated a normal lag phase early on, followed by the exponential phase after day 2. Samples 4A, 9A, and 26A seemed to enter the exponential growth phase right after the transfer, before entering the stationary phase. We again assume, the initial growth is due to the glucose residue from the earlier media. The slow growth indicates the adaptation to use xylose as the carbon source. The rest of the samples exhibit slow growth.

In MMX_B, all samples displayed growth, with sample 1B, 12B & 21B entering exponential phase early in the culture. Samples 3B, 18B and 23B exhibit exponential growth at day 1 and 15B continued to grow progressively till day 4. At this stage, a relationship can be seen between the success of yeasts in basic media able to assimilate xylose better than the ones in acidic media.

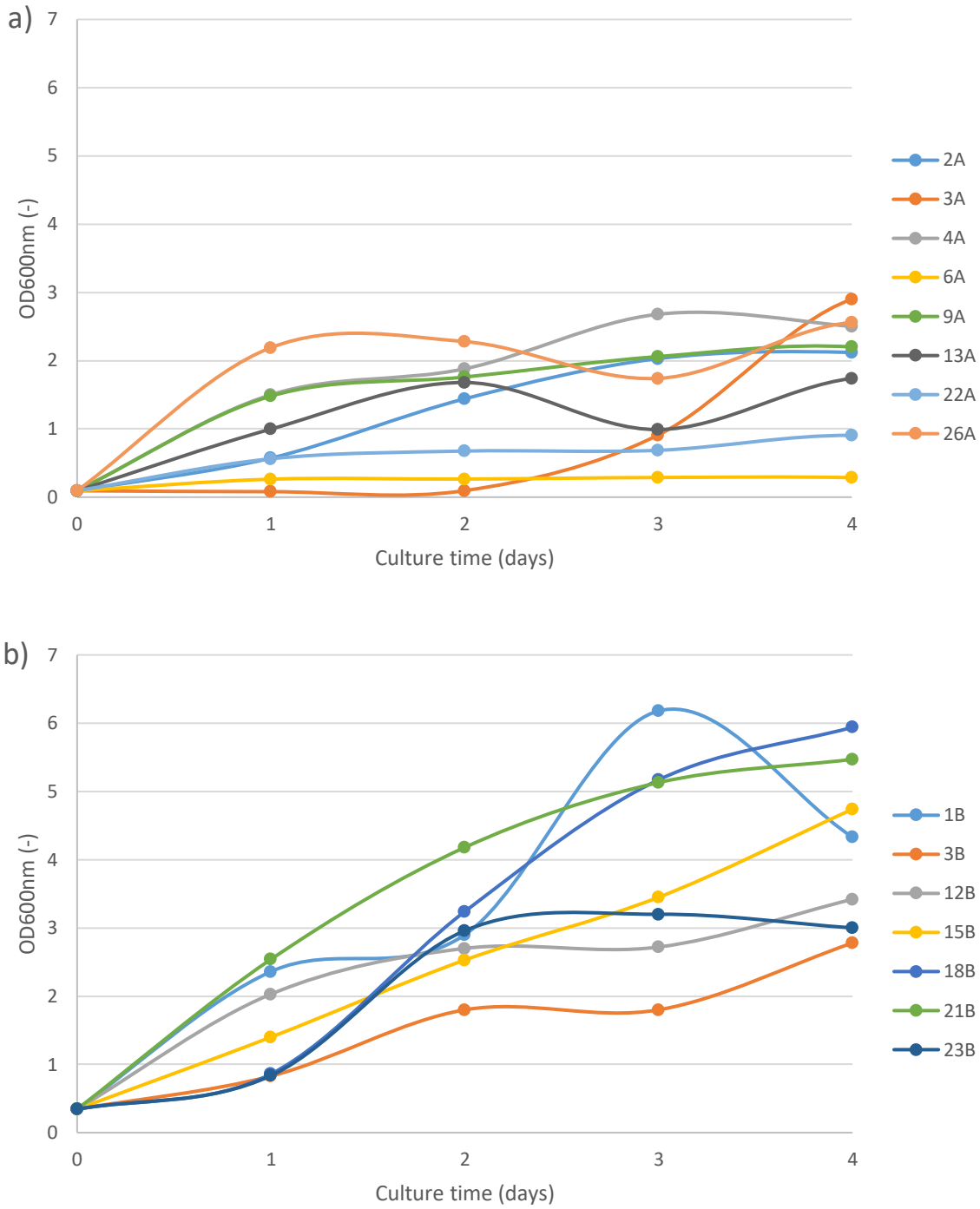
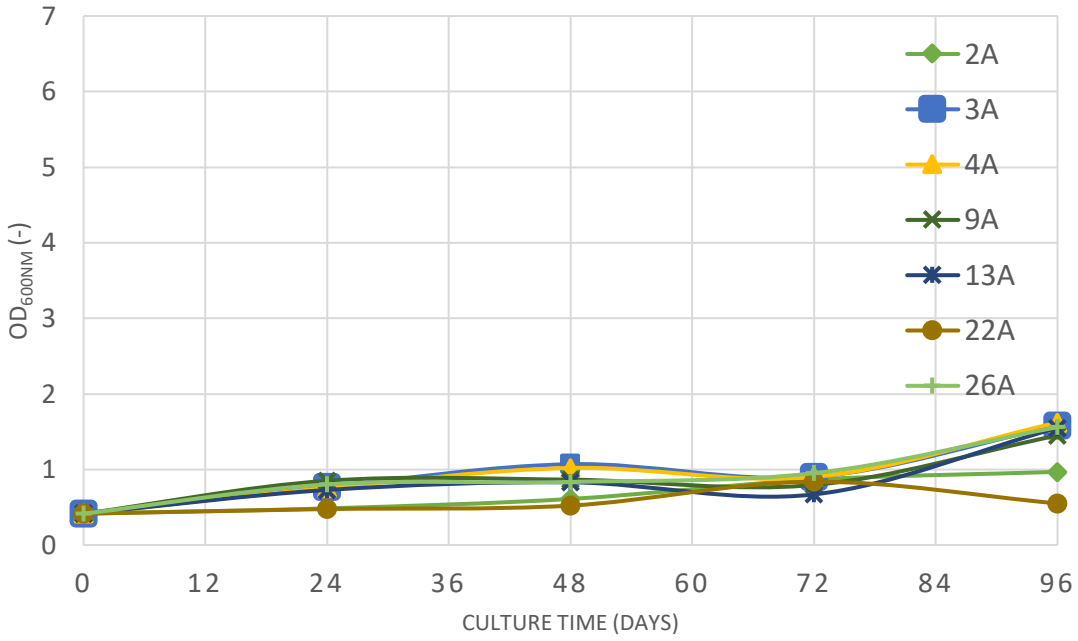


Figure 2.6: a) the growth of the unknown cultures that had survived the inhibitor stage of the screening on minimal media with xylose at acidic pH (MMX_A). b) The growth of cultures on minimal media with xylose at basic pH (MMX_B).

a)



b)

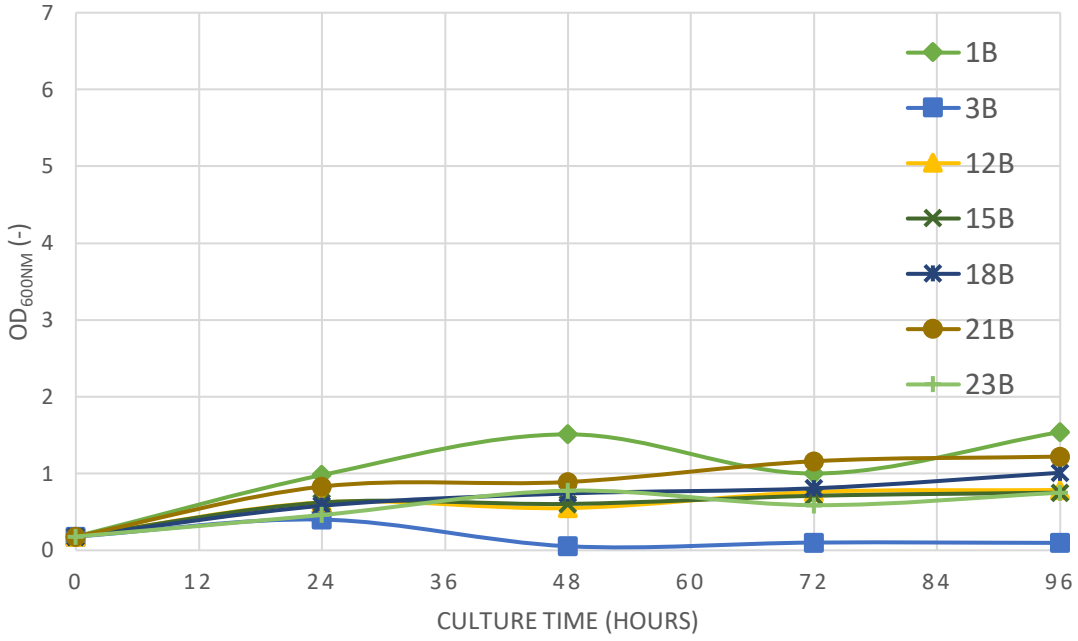


Figure 2.7: Growth of the unidentified cultures that had survived the previous three stages on minimal media with additional cellobiose (15 g/L) and arabinose (15 g/L) in a) acidic conditions and b) basic conditions.

Stage 4, growth of unknown biota on MMAC

Finally, the surviving yeasts from the xylose stage were cultured in minimal media with arabinose and cellobiose as the carbon sources (MMAC). Under these conditions the yeast grew slightly at a slow rate as shown in the lower O.D. readings (fig. 2.7). Many surviving strains mostly showed no growth. This must be due to the inability to metabolise the two sugars present. Arabinose and cellubiose need a cocktail of enzymes to catabolise them. While cellubiose only needs cellubiase to be utilised, arabinose needs D-arabitol dehydrogenase (ArDH) and ribulokinase to then enter the pentose phosphate pathway [111]. The growth is much slower than in xylose.

A few yeasts such as *Pichia stipitis, Candida shehatae* and *Pachysolen tannophilus* have shown to be able to use xylose competently [61]. Dien et al, 1996 have discovered that these species are adept to ferment arabinose: *Candida auringiensis, Candida succiphila, Ambrosiozyma monospora* [112]. A yeast known to be able to utilise cellobiase is *Candida queiroziae*, a species of yeast found in Brazil [113]. In the depolymerisation of lignocelluose, arabinose & cellobiose would only be produced in small amounts compared to glucose and xylose. Therefore, this insignificant production of such sugars, lessens the dependence on them as the major carbon sources.

2.3.5.2 Second Collection

Stage 1, growth of unknown biota on MML

25 tubes of samples were collected aeseptically from the the different locations in the compound of the University of Bath, its vicinity and 1 location in the outskirts of the city of Bath (Lansdowne). As done previously, each sample was rinsed carefully with Ringer's solution. The Ringer's solution with the captured microbiota was then transferred directly to the MML broth.

Each sample was then analysed at the endpoint, day (4) for each stage for the optical density to observe the growth of the culture (Table 2.5 a & b). Both acidic & basic conditions show excellent growth at the end of this stage of the experiment. In the acidic medium, only UB04A sample did not propagate at all. While all of the flasks for the basic medium exhibit substantial growth, the collection from the West Wood residential showed a sightly higher increase than the rest.

Stage 2, growth of unknown biota on MMI

Only 1 sample was eliminated from the acidic collection, and the rest were inoculated to the next stage including the inhibitors. This is the harshest phase and a test on the resilience of the strains in the samples. The endpoint O.D. reading was only taken at Day 7, so to permit strains to adapt to the acidity, basic & high inhibitors.

It has been demonstrated in this collection that the samples grown in MMI_A has little growth in the majority of them. Only 7 samples (UMMC1A, UBMC3A, UBMC5A, UBGC3A, LD02A AND WW01A) showed increase in O.D., with UBMC1A showing very little growth at day 7.

However, surprisingly, in the basic media, the opposite was true. High numbers of yeast demonstrated substantial growth. Only WW04B showed low significantly growth, while several others showed only three to six-fold of growth.

It would appear that the basic conditions stopped the mostly acid inhibitors from inhibiting growth of the organisms. The basic state does not provide an optimal condition for the action of the inhibitors to take place allowing the strains in the samples to exploit the nutrients available and adapt to the higher pH to grow in the longer period.

(a) Location of sampling (MML_A)	OD_{600nm} (-)	
	Day 0	Day 4
University of Bath Sports Training Centre 1 (UB01A)	1.230	19.50
University of Bath Sports Training Centre 2 (UB02A)	1.080	24.30
University of Bath Sports Training Centre 3 (UB03A)	1.070	9.510
University of Bath Sports Training Centre 4 (UB04A)	1.280	0.735
University of Bath Sports Training Centre 5 (UB05A)	1.200	7.670
University of Bath Sports Training Centre 6 (UB06A)	1.160	12.28
University Medical Centre area 1 (UBMC1A)	1.280	19.82
University Medical Centre area 2 (UBMC2A)	1.300	18.08
University Medical Centre area 3 (UBMC3A)	3.650	18.38
University Medical Centre area 4 (UBMC4A)	1.290	10.10
University Medical Centre area 5 (UBMC5A)	1.220	15.30
University Medical Centre area 6 (UBMC6A)	1.190	33.90
Roadside towards University Golf Club 1 (UBGC1A)	1.370	24.40
Roadside towards University Golf Club 2 (UBGC2A)	1.390	14.02
Roadside towards University Golf Club 3 (UBGC3A)	1.050	59.60
Roadside towards University Golf Club 4 (UBGC4A)	1.060	45.60
Lansdowne Race Course Carpark 1 (LD01A)	1.320	49.80
Lansdowne Race Course Carpark 2 (LD02A)	2.850	31.60
Behind Westwood Residence U. of Bath 1 (WW01A)	1.130	29.90
Behind Westwood Residence U. of Bath 2 (WW02A)	1.760	21.20
Behind Westwood Residence U. of Bath 3 (WW03A)	0.832	45.60
Behind Westwood Residence U. of Bath 4 (WW04A)	0.991	32.70
Behind Westwood Residence U. of Bath 5 (WW05A)	0.798	45.60
Behind Westwood Residence U. of Bath 6 (WW06A)	0.892	19.00
Behind Westwood Residence U. of Bath 7 (WW07A)	0.950	38.40

(b) Location of sampling (MML_B)	OD_{600nm} (-)	
	Day 0	Day 4
University of Bath Sports Training Centre 1 (UB01B)	1.180	30.20
University of Bath Sports Training Centre 2 (UB02B)	1.160	14.98
University of Bath Sports Training Centre 3 (UB03B)	1.120	2.50
University of Bath Sports Training Centre 4 (UB04B)	1.360	25.00
University of Bath Sports Training Centre 5 (UB05B)	0.467	31.20
University of Bath Sports Training Centre 6 (UB06B)	1.230	44.90
University Medical Centre area 1 (UBMC1B)	1.460	13.30
University Medical Centre area 2 (UBMC2B)	1.280	22.40
University Medical Centre area 3 (UBMC3B)	1.350	23.20
University Medical Centre area 4 (UBMC4B)	1.180	25.10
University Medical Centre area 5 (UBMC5B)	1.420	43.60
University Medical Centre area 6 (UBMC6B)	1.330	41.60
Roadside towards University Golf Club 1 (UBGC1B)	1.700	20.40
Roadside towards University Golf Club 2 (UBGC2B)	1.270	21.00
Roadside towards University Golf Club 3 (UBGC3B)	1.190	53.80
Roadside towards University Golf Club 4 (UBGC4B)	1.260	43.60
Lansdowne Race Course Carpark 1 (LD01B)	1.830	51.10
Lansdowne Race Course Carpark 2 (LD02B)	3.290	37.00
Behind Westwood Residence U. of Bath 1 (WW01B)	2.830	6.70
Behind Westwood Residence U. of Bath 2 (WW02B)	1.670	6.84
Behind Westwood Residence U. of Bath 3 (WW03B)	1.200	7.22
Behind Westwood Residence U. of Bath 4 (WW04B)	1.400	8.55
Behind Westwood Residence U. of Bath 5 (WW05B)	1.800	3.91
Behind Westwood Residence U. of Bath 6 (WW06B)	1.300	3.88
Behind Westwood Residence U. of Bath 7 (WW07B)	1.400	5.01

Table 2.5: (a) Growth of unidentified cultures sampled from the University of Bath & Bath area, in an acidic minimal medium (MML_A) containing glucose, lysine & antibiotics. The starting pH of the cultures grown in pH 4.8 at day 0, and endpoint (OD_{600nm}) reading at day 4.

(b) Growth of unidentified cultures sampled from the University of Bath & Bath area, in a basic minimal medium (MML_B) containing glucose, lysine & antibiotics. The starting pH of the cultures grown in pH 9.0 at day 0, and endpoint (OD_{600nm}) reading at day 4.

(a)	Location of sampling (MMI_A)	OD_{600nm} (-)	
		Day 0	Day 7
University of Bath Sports Training Centre 1 (UB01A)		2.090	1.540
University of Bath Sports Training Centre 2 (UB02A)		2.410	2.000
University of Bath Sports Training Centre 3 (UB03A)		1.360	0.885
University of Bath Sports Training Centre 4 (UB04A)		0.621	0.618
University of Bath Sports Training Centre 5 (UB05A)		0.900	0.751
University of Bath Sports Training Centre 6 (UB06A)		1.100	0.822
University Medical Centre area 1 (UBMC1A)		1.640	1.650
University Medical Centre area 2 (UBMC2A)		0.782	0.708
University Medical Centre area 3 (UBMC3A)		0.235	1.390
University Medical Centre area 4 (UBMC4A)		1.660	1.160
University Medical Centre area 5 (UBMC5A)		1.490	1.660
University Medical Centre area 6 (UBMC6A)		2.910	1.220
Roadside towards University Golf Club 1 (UBGC1A)		0.240	1.320
Roadside towards University Golf Club 2 (UBGC2A)		1.810	1.230
Roadside towards University Golf Club 3 (UBGC3A)		4.940	5.730
Roadside towards University Golf Club 4 (UBGC4A)		3.770	3.240
Lansdowne Race Course Carpark 1 (LD01A)		3.230	2.74
Lansdowne Race Course Carpark 2 (LD02A)		1.520	7.39
Behind Westwood Residence U. of Bath 1 (WW01A)		2.150	2.930
Behind Westwood Residence U. of Bath 2 (WW02A)		1.330	0.777
Behind Westwood Residence U. of Bath 3 (WW03A)		2.480	2.180
Behind Westwood Residence U. of Bath 4 (WW04A)		2.230	1.790
Behind Westwood Residence U. of Bath 5 (WW05A)		3.180	2.330
Behind Westwood Residence U. of Bath 6 (WW06A)		2.230	0.990
Behind Westwood Residence U. of Bath 7 (WW07A)		3.180	2.140

(b)	Location of sampling (MMI_B)	OD_{600nm} (-)	
		Day 0	Day 7
University of Bath Sports Training Centre 1 (UB01B)		3.060	47.80
University of Bath Sports Training Centre 2 (UB02B)		2.770	3.560
University of Bath Sports Training Centre 3 (UB03B)		1.870	6.750
University of Bath Sports Training Centre 4 (UB04B)		6.330	49.70
University of Bath Sports Training Centre 5 (UB05B)		4.170	5.480
University of Bath Sports Training Centre 6 (UB06B)		4.670	5.640
University Medical Centre area 1 (UBMC1B)		2.870	37.60
University Medical Centre area 2 (UBMC2B)		1.930	5.850
University Medical Centre area 3 (UBMC3B)		2.430	5.320
University Medical Centre area 4 (UBMC4B)		2.610	44.40
University Medical Centre area 5 (UBMC5B)		4.040	40.50
University Medical Centre area 6 (UBMC6B)		4.600	45.30
Roadside towards University Golf Club 1 (UBGC1B)		2.580	6.950
Roadside towards University Golf Club 2 (UBGC2B)		2.050	6.070
Roadside towards University Golf Club 3 (UBGC3B)		3.700	9.140
Roadside towards University Golf Club 4 (UBGC4B)		5.190	20.30
Lansdowne Race Course Carpark 1 (LD01B)		5.70	53.5
Lansdowne Race Course Carpark 2 (LD02B)		5.63	72.2
Behind Westwood Residence U. of Bath 1 (WW01B)		5.48	51.5
Behind Westwood Residence U. of Bath 2 (WW02B)		3.77	10.9
Behind Westwood Residence U. of Bath 3 (WW03B)		5.48	49.4
Behind Westwood Residence U. of Bath 4 (WW04B)		3.59	1.80
Behind Westwood Residence U. of Bath 5 (WW05B)		4.74	62.1
Behind Westwood Residence U. of Bath 6 (WW06B)		3.92	35.5
Behind Westwood Residence U. of Bath 7 (WW07B)		5.04	46.9

Table 2.6: (a) Growth of of unidentified cultures sampled from the University of Bath & Bath area, in acidic minimal medium (MMI_A) with glucose (30 g/L) and inhibitors (furfural (10mM), acetic acid (60mM), formic acid (60mM, levullinic acid (60mM, 5-HMF (10mM). The starting pH of the cultures grown in pH 4.8 at day 0, and endpoint (OD_{600nm}) reading at day 4.

(b) Growth of unidentified cultures sampled from the University of Bath & Bath area, in acidic minimal medium (MMI_B) with glucose (30 g/L) and inhibitors (furfural (10mM), acetic acid (60mM), formic acid (60mM, levullinic acid (60mM, 5-HMF (10mM). The starting pH of the cultures grown in pH 9.0 at day 0, and endpoint (OD_{600nm}) reading at day 4.

Stage 3, growth of unknown biota on MMX

In the next step, the surviving samples were transferred into MMX (acid and basic) media. At the endpoint of day 4, half the samples showed significant growth, while only 1 flask (UBMC5A) showed slow growth in the acidic medium. This was surprising due to the relative lack of xylose assimilating yeast from the previous study.

In the basic set, as they have shown high growth rate in the MMI_B medium, the inoculum which provide them with glucose supply helped them to grow and persist continuing growing in the Xylose medium. Only 3 samples showed negative O.D. changes and the rest were taken to the next stage of filtering (fig 2.8).

Stage 4, growth of unknown biota on MMAC

In the last stage, the inoculum from the earlier stage was transferred in to the MMAC (acid and basic) media. Arabinose and cellobiose are not commonly available sugars in nature. In the acidic medium, all 5 samples did show significant growth. This can be a good indication of the adaptability of the strains in the collection as readily adaptable to different conditions incurred.

In the $MMAC_B$, 9 samples did not grow while the rest showed slightly slow growth which is expected due to the less preferred sugars present.

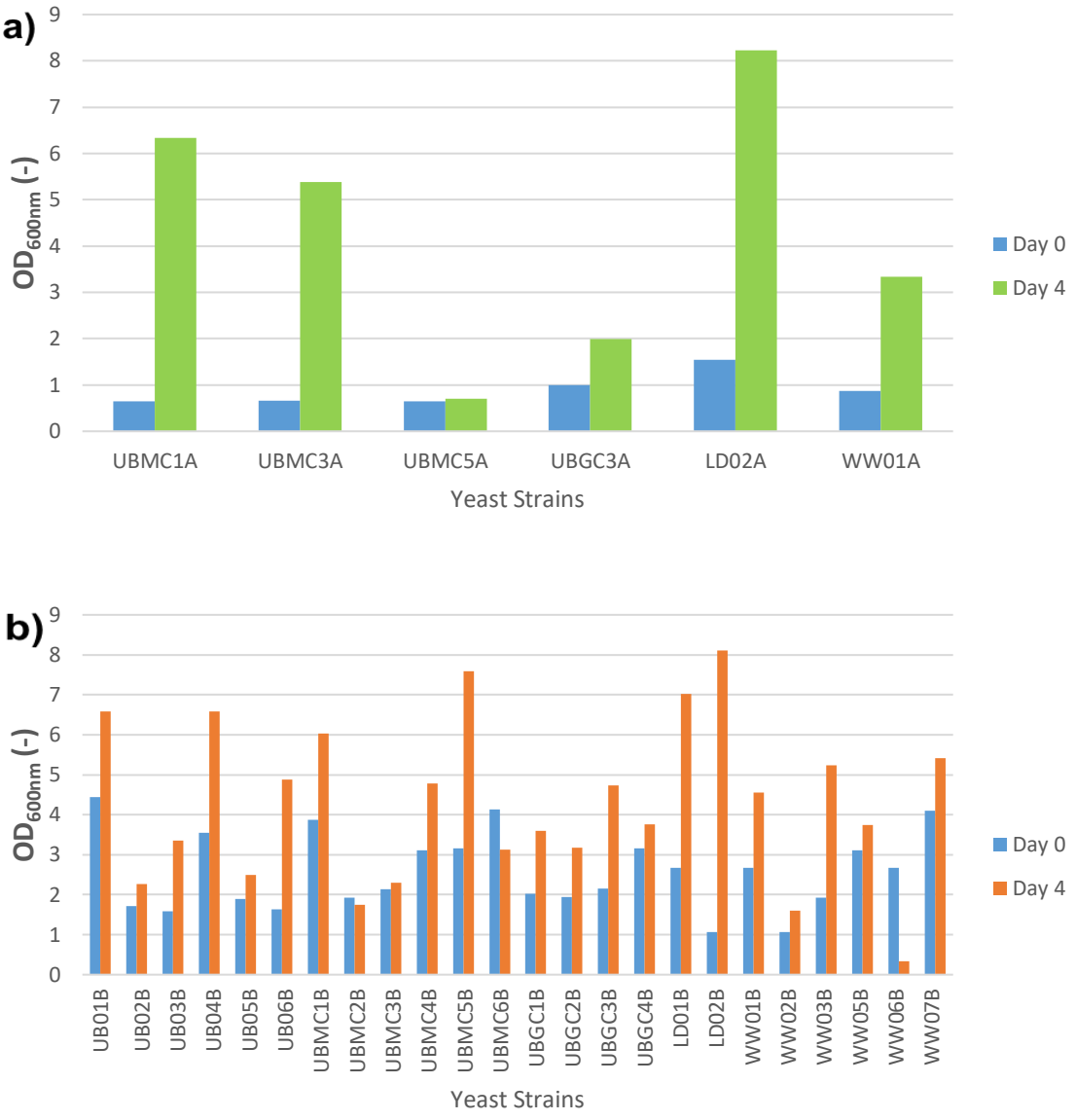


Figure 2.8: (a) Growth of unidentified cultures sampled from the University of Bath & Bath area, in and acidic (MMX_B) minimal medium with xylose (30 g/L). The starting pH of the cultures grown in pH 4.8 at day 0, and endpoint (OD_{600nm}) reading at day 4.

(b) Growth of unidentified cultures sampled from the University of Bath & Bath area, in and basic (MMX_B) minimal medium with xylose (30 g/L). The starting pH of the cultures grown in pH 9.0 at day 0, and endpoint (OD_{600nm}) reading at day 4.

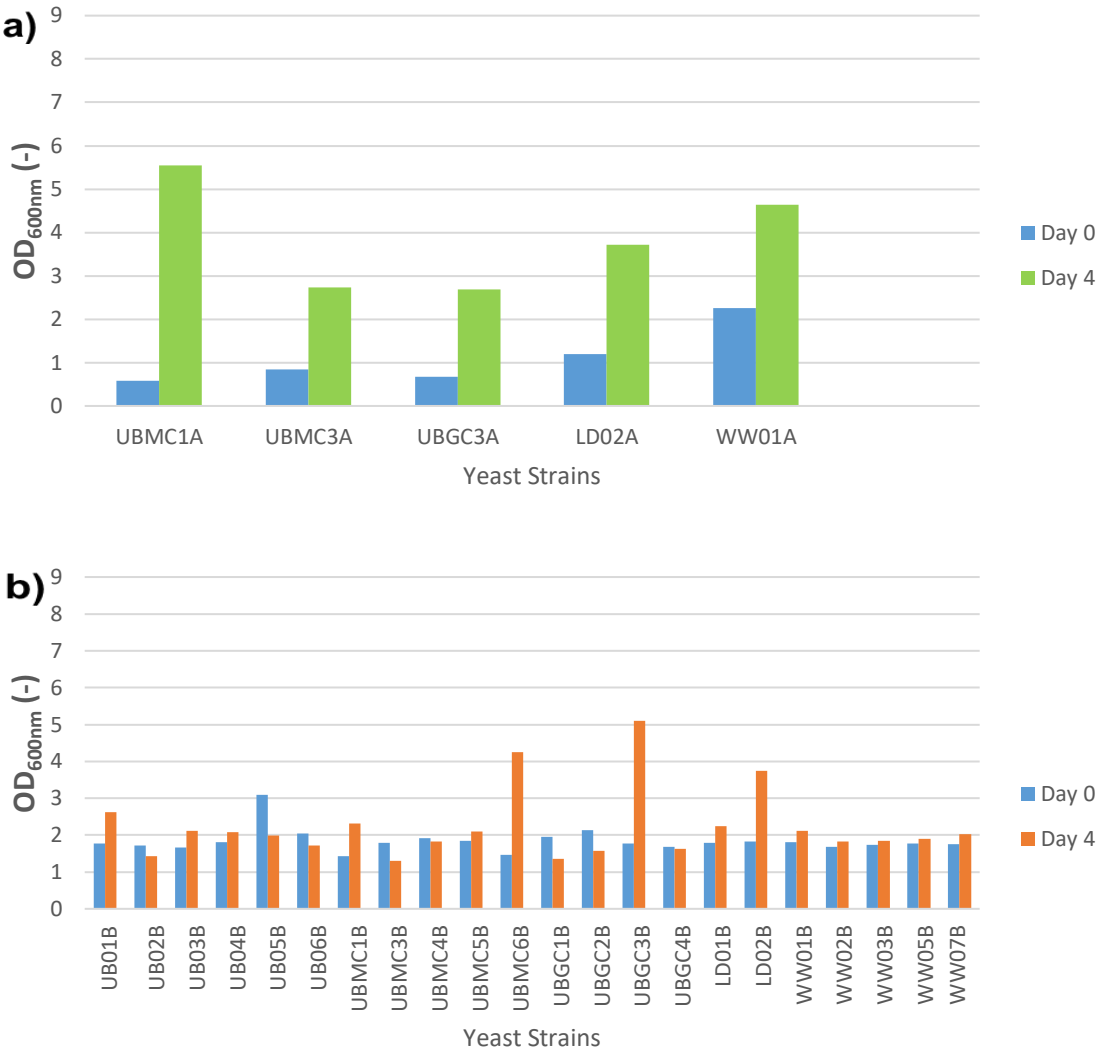


Figure 2.9: (a) Growth of unidentified cultures sampled from the University of Bath & Bath area, in minimal medium with arabinose, 15 g/l and cellobiose, 15 g/l. The starting pH of the cultures grown in pH 4.8 at day 0, and endpoint (OD_{600nm}) reading at day 4. (b) Growth of unidentified cultures sampled from the University of Bath & Bath area, in minimal medium with arabinose, 15 g/l and cellobiose, 15 g/l. The starting pH of the cultures grown in pH 9.0 at day 0, and endpoint (OD_{600nm}) reading at day 4.

2.3.6 Identification of novel strains

2.3.6.1 First Collection

For the first set of collections, in identifying of the yeast, an identification kit, API® ID 32 strip from bioMérieux, Inc. was used. The microorganisms to be identified were first isolated and streaked onto an agar plate. One or a few colonies, aged between 24-48 hours, were then transferred into distilled water to meet a turbidity of 2 McFarland. This was then transferred to API® C Medium, which afterward was dispensed into each cupule of the strip. It was then incubated for 24-48 hours at 29°C.

This technique consists of a single-use plastic strip consisting of 32 wells. Each contains substrates for 29 assimilation tests (carbohydrates, organic acids, and amino acids), one susceptibility test (cycloheximide), and one colorimetric test (esculin). [145, 146]

Visual observation was used to determine growth either to be positive or negative based upon the presence or absence of turbidity in the wells. A manual interpretation was used to identify the yeast species, observing if there was any reaction in the 32 cupules. A positive reaction was given as an increase in the turbidity of the cupules. Only two out of the 14 species could be identified. Although the two colonies were from a slow growing sample (in MMAC), these were the only successful identification using the API® ID 32 kit. This may be caused by the yeasts sampled not being present in the listed identifiable yeasts in the kit. Therefore, due to these difficulties, a n alternate method was used to identify the species found in this first collection and the subsequent bioprospecting collections.

The two yeasts identified are:

a) 23B- *Trichosporon mucoides*; b) 22AIII- *Candida valida*

Neither species has been investigated previously for the ability to produce suitable bioproducts. However, the genus *Candida* has been historically known to be oleaginous since the commercialization of yeast as protein producers that lead to the extensive studies on *Candida utilis* as a producer of nutritional oils [147]. Many other species of *Candida* are also proven to be lipid producing yeasts such as *Candida*

guilliermondii[148], *Candida freyschussii* [149], *Candida cleridarum* [150], *Candida diddensiae* [151], and *Candida tropicalis* [152].

2.3.6.2 Second Collection

To provide a more suitable identification, a molecular technique was used for the samples collected from different locations in the Bath area. In addition, in using molecular techniques such as DNA sequencing, a database can be created and analysed.

As yeast is under the fungi kingdom as the second largest group, a "barcode" in identification of species has been developed over 15-20 years. Barcodes are basically short, standardised DNA regions used to identify biological materials [114]. In this project, we selected Internal Transcribed Spacer Subunit (of nuclear ribosomal DNA genes) as the first concept to be utilised in the identification process. This has been the favourable DNA barcoding marker "for the identification of single taxa and mixed environmental templates" and proposed by majority of mycologists as the principal barcoding marker for fungi [115]. This is even more substantial as the database for such sequences are overtime deposited into public accessible portals with built-in applications for data manipulations and data building as in the National Center for Biotechnology Information (NCBI).

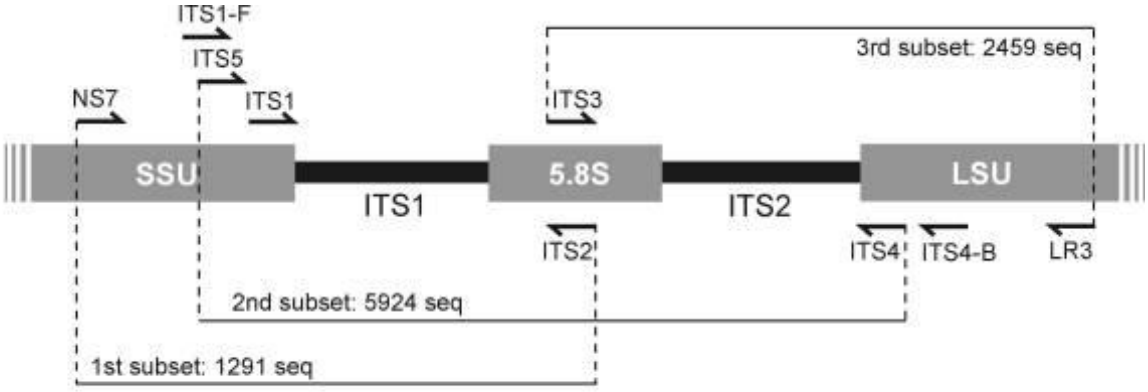


Figure 2.10: ITS genes on the nuclear ribosomal DNA. Extracted from Bellmain et al. 2010 [115]

In this study the ITS1-F and LR3-R primers were used to amplify DNA from the unknown taxas. A standard PCR technique was used on all of the individual strains isolated from the WL plates into individual YPD agar plates. Cells from a single colony were gently taken using an inoculation loop and mixed with the primers above and a standard Master Mix which includes Taq polymerase for PCR to be catalyzed. PCR conditions were optimized for each primer combination, but the general reaction protocol was as follows: initial denaturation of 94°C for 5 min followed by 25 cycles of 1 min (94°C); 1 min (55°C); 2 min (72°C) followed by a final extension of 72°C for 7 min. PCR products were visualized on 1.5% agarose gels.

Then, the successful PCR products as indicated by the bands shown (fig. 2.7), were prepared onto the sequencing tubes and sent off to Eurofin Genomics for DNA sequencing.

When the data arrived, the DNA sequences were uploaded into the Portal BLAST (Basic Local Alignment Search Tool) from the NCBI website. To use it, a researcher submits to the algorithm a sequence of interest. The sequence can be DNA, RNA, or an amino acid chain. The algorithm will then compare the sequence the user submitted with the sequences in its database, and tell the user which database sequence most closely matches the user-submitted sequence.

In the first trial only 10 strains were successfully identified, as not all strains went through the PCR successfully. We had to opt using the Bead Beater® procedure to extract the DNA from the inoculated cells. After which only can the DNA be positively visualized using the Gel Electrophoresis method. Again, the same preparations were made to the cells to be sequenced and then aligned on the BLAST applications. In this batch we maintained using the ITS-F & LR-3 primers combination in the PCR steps. This was very successful, and finally, all of the strains were identified, using the ITS-F & LR-3 primers combination except for one using the ITS1-F and ITS4-R combinations for LD02A (table 2.7).

Table 2.7: Strains identified from PCR-DNA sequencing method using ITS primers

No.	Label	*Species*	Strains From NCBI database	BLAST Accession No.
1	UBMC1	*Metschnikowia aff. chrysoperlae*	98% P34A005	FR07618459 (784 bases)
2	WW01BII	*Metschnikowia aff. chrysoperlae*	98% NRRL Y-6259	FR07618472 (901 bases)
3	UBMC6B	*Metschnikowia aff. chrysoperlae*	99% P44A006	FR07618462 (745 bases)
4	WW05B	*Metschnikowia aff. chrysoperlae*	96% P34A005	FR07618475 (923 bases)
5	WW02B	*Metschnikowia pulcherrima*	98% NRRL Y-7111	FR07618473 (861 bases)
6	UBGC3AI	*Metschnikowia pulcherrima*	98% XY103	FR07618464 (815 bases)
7	UBGC3AII	*Metschnikowia pulcherrima*	UNIDENTIFIED	FR07618465 (749 bases)
8	UBGC1	*Candida friedrichii*	99% NBRC 10277	FR07618463 (890 bases)
9	UB04BII	*Candida oleophila*	99% P40C007	FR07618467 (925 bases)
10	LD01BII	*Meyerozyma guilliermondii*	99% KAML05	FR07618468 (998 bases)
11	LD02A	*Wickerhamomyces anomalus*	100% P42B001	FR07618498 (602 bases)
12	UB04BI	*Rhodotorula mucilaginosa*	93% DY115-21-3-Y25	FR07618506 (998 bases)

From this results, we have more than half are of the collection of M. aff. Chrysoperlae and M. pulcherrima which are very closely related phylogenetically. This will be proven later in a multi locus study of the seven strains. M. pulcherrima (anamorph Candida pulcherrima) has been known as oleaginous by Pan et al in 2009 [162]. The genuses Candida, Rhodotorula and Wickerhamomyces have been known to be lipid-producing too [163].

Table 2.8 exhibit the DNA sequences of the strains used in the identification using the BLAST application on NCBI portal.

Table 2.8: DNA Sequences of the identified strains

1. UBMC1 >*Metschnikowia aff. chrysoperlae* >FR07618459 (784 bases)
TTCCCCGCCGCCCCGATGCTGGCCCAGTTAAGTGTCTGCTTGCAAGCCCTTCCCTTTCAACAATTTCACG TACTTTTTCACTCTCTTTTCAAAGTGCTTTTCATCTTTCCATCACTGTACTTGTTCGCTATCGGTCTCTC GCCGGTATTTAGCTTTAGATGGAATTTACCACCCACTTAGAGCTGCATTCCCAAACAACTCGACTCTTGG AGATCTGGGATGAGGGCGTTAAAGGGGTTCACGGGGCTGTCACCCTCTGTGGCGCCACTTTCCAGTGGAC TTAACCCCTGCCGGCCGGACCCAAATCTCTTCAAATTACAATTCCCGGGGGATTTCAAATTTGAGCTTTT GCCGCTTCACTCGCCGTTACTGAGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAAG TTCAGCGGGTAGTCTTACGAGGGTGAGGAAAAAGAATGGGGCTAAAACTTATTCTAGCGCCGTTGATATT AGGCCGAAGCAGGACCAAACCGGAGGTTTGAGAGTAAATATCGCTCACCCACGCATGCCCTGGGGAATAC CCCGGGGCGCAATGTGCGTTCAAAGATTCAATGATTCACGTCTGCAAGTCATATTACGTATCGCAATTCG CTGCGTTCTTCATCGATGCGAGAACCAAGAGATCCGTTGTTGAAAGTTTTTTAATTGATGTATTTTGAGG ATAAAGAGTTCAGAGTGTTGTTGTCCTAAAAGTGTGTAAAAATATATATCTAATGATCCTTCCGCAGGTT CACCTACGGGAAAC
2. WW01BII >*Metschnikowia aff. Chrysoperlae* >FR07618472 (901 bases)
CGAGGATGGCGTGTCGATCATGCATAGATGAGCATCAGGGTGGTGAGGGAGTTAGGGACCGGGGTTATAA TCCTCGAAAGGTACATTCCCCGGTGGTTTTGTTCCCCGCCGCCCCGATGCTGGCCCAGTTAAGTGTCTGC TTGCAAGCCCTTCCCTTTCAACAATTTCACGTACTTTTTCACTCTCTTTTCAAAGTGCTTTTCATCTTTC CATCACTGTACTTGTTCGCTATCGGTCTCTCGCCGGTATTTAGCTTTAGATGGAATTTACCACCCACTTA GAGCTGCATTCCCAAACAACTCGACTCTTGGAGATCTGGGATGAGGGCGTTGAAGGGGTTTACGGGGCTG TCACCCTCTGTGGCGCCACTTTCCAGTGGACTTAACCCCTGCCGGCCGGACCCAAATCTCTTCAAATTAC AATTCCCGGGGGATTTCAAATTTGAGCTTTTGCCGCTTCACTCGCCGTTACTGAGGCAATCCCTGTTGGT TTCTTTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTAGTCTTACGAGGGTGAGGAAGAAGAATGG GGCTAAAACTTATTCTAGCGCCGTTGATATTAGGCCGAAGCATGACCAAACCGGAGGTTTGAGAGTAAAT ATCGCTCACCCACGCATGCCCTGGGGAATACCCCGGGGCGCAATGTGCGTTCAAAGATTCAATGATTCAC GTCTGCAAGTCATATTACGTATCGCAATTCGCTGCGTTCTTCATCGATGCGAGAACCAAGAGATCCGTTG TTGAAAGTTTTTTTGAAAAAAATATTGACGGTTAAGATTTAGAGTTTGTGCCTAAAAGTGTGTATATCAT AATATGATCTCGCAGTCACTACGAAACTGTACGACTTTACTCCACATTGTGTGCCCGGGGA
3. UBMC6B >*Metschnikowia aff. Chrysoperlae* >FR07618462 (745 bases)
TAAGGACCGGGGTTATAATCCTCGAAAGGTACATTCCCCGGTGGTTTTGCTCCCCGCCGCCCCGATGCTG GCCCAGTTAAGTGTCTGCTTGCAAGCCCTTCCCTTTCAACAATTTCACGTACTTTTTCACTCTCTTTTCA AAGTGCTTTTCATCTTTCCATCACTGTACTTGTTCGCTATCGGTCTCTCGCCGGTATTTAGCTTTAGATG GAATTTACCACCCACTTAGAGCTGCATTCCCAAACAACTCGACTCTTGGAGATCTGGGATGAAGGCGTTG AAGGGGTTCACGGGGCTGTCACCCTCTGTGGCGCCACTTTCCAGTGGACTTAACCCCTGCCGGCCGGACC CAAATCTCTTCAAATTACAATTCCCGGGGGATTTCAAATTTGAGCTTTTGCCGCTTCACTCGCCGTTACT GAGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTAGTCTTACGAG GGTGAGGAAAAAGGATGGGGCTAAAACTTATTCTAGCGCCGTTGATATTAGGCCGAAGCAGGACCAAACC GGAGGTTTGAGAGTAAATATCGCTCACCCACGCATGCCCTGGGGAATACCCCGGGGCGCAATGTGCGTTC AAAGATTCAATGATTCACGTCTGCAAGTCATATTACGTATCGCAATTCGCTGCGTTCTTCATCGATGCGA GAACCAAGAGATCCGTTGTTGAAAGTTTTTTAATTGAGTAATTGA
4. WW05B >*Metschnikowia aff. Chrysoperlae* >FR07618475 (923 bases)
CGGGGAATGCCACTGGTCCGCTGATGACCATGAGGGCACAAGGAAATTGAGACCGGGGTTATAATCCTCG AAAGGTACATTCCCCGGTGGTTTTGTTTCCCGCCGCCCCGATGCTGGCCCAGTTAAGTGTCTGCTTGCAA GCCCTTCCCTTTCAACAATTTCACGTACTTTTTCACTCTCTTTTCAAAGTGCTTTTCATCTTTCCATCAC TGTACTTGTTCGCTATCGGTCTCTCGCCGGTATTTAGCTTTAGATGGAATTTACCACCCACTTAGAGCTG CATTCCCAAACAACTCGACTCTTGGAGATCTGGGGTAGAGGCTTTAGAGGGGCTCACGGGGCTGTCACCC TCTGTGGCGCCACTTTCCAGTGGACTTAACCCCTGCCGGCCGGACCCAAATCTCTTCAAATTACAATTCC CGGGGGATTTCAAATTTGAGCTTTTGCCGCTTCACTCGCCGTTACTGAGGCAATCCCTGTTGGTTTCTTT TCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTAGTCTTACGAGGGTGAGGAAAAAGAATGGGGCTAA AACTTATTCTAGCGCCGTTGATATTAGGCCGAAGCAGGACCAAACCGGAGGTTTGAGAGTAAATATCGCT CACCCACGCATGCCCTGGGGAATACCCCGGGGCGCAATGTGCGTTCAAAGATTCAATGATTCACGTCTGC AAGTCATATTACGTATCGCAATTCGCTGCGTTCTTCATCGATGCGAGAACCAAGAGATCCGTTGTTGAAA GTTTTTTGTATTTGATGTTATTTGAAGGAGTATAGATCTAGTAGTTTGTTGCCTAAAAGTGTGTAATAAT AATTATTAATGATCCTTTCCGCAGGTTCACCTACGGAAACCCTTGTTACGACTTTTACTTCTCATTAAAA ATGGAACCAAAGG

5. WW02B >*Metschnikowia pulcherrima* >FR07618473 (861 bases)
GAGGGCAGTATGGGAGTAAGGGCCGGGGTTATAATCCTCGAAAGGTACATTCCCCAGTGGTTTTGCTCCC CGCCGCCCCGATGCTGGCCCAGTTAAGTGTCTGCTTGCAAGCCCTTCCCTTTCAACAATTTCACGTACTT TTTCACTCTCTTTTCAAAGTGCTTTTCATCTTTCCATCACTGTACTTGTTCGCTATCGGTCTCTCGCCGG TATTTAGCTTTAGATGGAATTTACCACCCACTTAGAGCTGCATTCCCAAACAACTCGACTCTTGGAGATC TGGGATGAAGGCGTTGAGGGGGTTCACGGGGCTGTCACCCTCTGTGGCGCCACTTTCCAGTGGACTTAAC CCCTGCCGGCCGGACCCAAATCTCTTCAAATTACAATTCCCGGGGGATTTCAAATTTGAGCTTTTGCCGC TTCACTCGCCGTTACTGAGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAAGTTCAG CGGGTAGTCTTACGAGGGTGAGGAAAAAGGATGGGGCTAAAACTTATTCTAGCGCCGTTGATATTAGGCC GAAGCAGGACCAAACCAGAGGTTTGAGAGTAAATATCGCTCACCCACGCATGCCCTGGGGAATACCCCGG GGCGCAATGTGCGTTCAAAGATTCAATGATTCACGTCTGCAAGTCATATTACGTATCGCAATTCGCTGCG TTCTTCATCGATGCGAGAACCAAGAGATCCGTTGTTGAAAGTTTTTTGATAAAGTTATTGAGGGTTAAGA TTTAGAGTTTGTGCCTAAAAGGGTGTAAAAACAATATTAATGATCCTTCCGCAGGTTCACCTACGGAAAC CTTGTTACGACTTTACTCCTT
6. UBGC3AI >*Metschnikowia pulcherrima* >FR07618464 (815 bases)
CACAAGGAAATTGAGACCGGGGTTATAATCCTCGAAAGGTACATTCCCCGGTGGTTTTGTTTCCCGCCGC CCCGATGCTGGCCCAGTTAAGTGTCTGCTTGCAAGCCCTTCCCTTTCAACAATTTCACGTACTTTTTCAC TCTCTTTTCAAAGTGCTTTTCATCTTTCCATCACTGTACTTGTTCGCTATCGGTCTCTCGCCGGTATTTA GCTTTAGATGGAATTTACCACCCACTTAGAGCTGCATTCCCAAACAACTCGACTCTTGGAGATCTGGGAT GAGGGCGTTGAGGGGGTTCACGGGGCTGTCACCCTCTGTGGCGCCACTTTCCAGTGGACTTAACCCCCGC CGGCCGGACCCAAATCTCTTCAAATTACAATTCCCGGGGGATTTCAAATTTGAGCTTTTGCCGCTTCACT CGCCGTTACTGAGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTA GTCTTACGAGGGTGAGGAAAAAGGCTGGGGCTAAAACTTATTCTAGCGCCGTTGATATTATGCCGAAGCA GGACCAAACCAGAGGTTTGAGAGTAAATATCGCTCACCCACGCATGCCCTGGGGAATACCCCGGGGCGCA ATGTGCGTTCAAAGATTCAATGATTCACGTCTGCAAGTCATATTACGTATCGCAATTCGCTGCGTTCTTC ATCGATGCGAGAACCAAGAGATCCGTTGTTGAAAGTTTTTTAATTAGTGTTATTGAAGAAATAAGATTCA GATGTTTTTTCCTAAAAGTGTGGTAATAATAATTTTTAATGATCC
7. UBGC3AII >*Metschnikowia pulcherrima* >FR07618465 (749 bases)
CATGGAAATTAGAGACCAGGGGTTATAATCCTCGAAAGGTACATTCCCCGGTGGTTTTGTTTCCCGCCGC CCCGATGCTGGCCCAGTTAAGTGTCTGCTTGCAAGCCCTTCCCTTTCAACAATTTCACGTACTTTTTCAC TCTCTTTTCAAAGTGCTTTTCATCTTTCCATCACTGTACTTGTTCGCTATCGGTCTCTCGCCGGTATTTA GCTTTAGATGGAATTTACCACCCACTTAGAGCTGCATTCCCAAACAACTCGACTCTTGGAGATCTGGGAT GAGGGCGTTGAGGGGGTTCACGGGGCTGTCACCCTCTGTGGCGCCACTTTCCAGTGGACTTAACCCCCGC CGGCCGGACCCAAATCTCTTCAAATTACAATTCCCGGGGGATTTCAAATTTGAGCTTTTGCCGCTTCACT CGCCGTTACTGAGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTA GTCTTACGAGGGTGAGGAAAAAGGCTGGGGCTAAAACTTATTCTAGCGCCGTTGATATTATGCCGAAGCA GGACCAAACCAGAGGTTTGAGAGTAAATATCGCTCACCCACGCATGCCCTGGGGAATACCCCGGGGCGCA ATGTGCGTTCAAAGATTCAATGATTCACGTCTGCAAGTCATATTACGTATCGCAATTCGCTGCGTTCTTC ATCGATGCGAGAACCAAGAGATCCGTTGTTGAAAGTTTTTTAATTAGTG
8. UBGC1 >Candida friedrichii >FR07618463 (890 bases)
GCGGGCCTCAGTCTAGATAGGCAGTATCGACGAAGTCTATAATACACCACCGAAGTAGTGTTACTTTCCA ACGCAATTATCCTGCCATCCAAACTGATGCTGGCCCGGAAAGCTCCAAAAATGGAAAATTCCAAGTCTAA TCTCAAGCCCTTCCCTTTCAACAATTTCACGTACTTTTTCACTCTCTTTTCAAAGTTCTTTTCATCTTTC CTTCACAGTACTTGTTCGCTATCGGTCTCTCGCCAATATTTAGCTTTAGATGGAATTTACCACCCACTTA GAGCTGCATTCCCAAACAACTCGACTCGTCGAAGGAACTTTACATGGAATTGGGCATCTCATCGCACGGG ATTCTCACCCTCTGTGACGTCCTGTTCCAAGGAACATAGACAAGAGCCAAAACCAAAGATACCTTCTTCA AATTACAACTCGGACACCGAAGGCGCCAGATTTCAAATTTGAGCTATTGCCGCTTCACTCGCCGCTACTA AGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAAGTTCAGCGGGTATTCTTACCTGA TTTGAGGTCAAACTTAGTTTGTTATTATAAGGCCGAGCCAATTACTAGAAATTTACTACCCCAGTCTTCA ACGAGTTGGATAAACCTAATACATTGAATAACCAGTTCGTACTATCCAGTACACTCCTGCCATTATATTT AAAGCAAACACTTAGTTGACTAAGAGTATCACTCAACACCAAACCCAAAGGTTTGAAGGAGAAATGACGC TCAAACAGGCATGCCCTTTGGAATACCAAAGGGCGCAATGTGCGTTCAAAGATTCGATGATTCACGAAAA TCTGCAATTCATATTACTTATCGCATTTCGCTGCGTTCTCATCGATGCGA

9. UB04BII >*Candida oleophila* >FR07618467 (925 bases)

CCAGCATCGGTCTAGATAGGCAGTATCAACAATATGTCTATAACACTTCACCGAAGCAAAGCTACATTCC
AATGTCATTATCCTACCATCCAAACCGATGCTGGCCCGGTAAGCTGTGAGGAACCAAATCCAAAAAGAAT
AAGGAACACAAAATACCAAGTCTGATCTCAAGCCCTTCCCTTTCAACAATTTCACGTACTTTTTCACTCT
CTTTTCAAAGTTCTTTTCATCTTTCCATCACTGTACTTGTTCGCTATCGGTCTCTCGCCAATATTTAGCT
TTAGATGGAATTTACCACCCACTTAGAGCTGCATTCCCAAACAACTCGACTCTTCGAAAGCACTTTACAT
AGAATTGGACATCTCATCGCACGGGATTCTCACCCTCTGTGACGTCCTGTTCCAAGGAACATAGACAAGA
GCCAATTCCAAAGTTACCTTCTTCAAATTACAACTCGAACACTGAAAGTGCTAGATTTCAAATTTGAGCT
TTTGCCGCTTCACTCGCCGTTACTAAGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTT
AAGTTCAGCGGGTAGTCCTACCTGATTTGAGGTCAAACTTTGTTTGTTGTTGTAAGGCCGAGCCTGTGTA
ATAAACAAATACTTTACTGGTCAACGAGTTGGATAAACCTAATACATTGAAAACAGTTCAGCACTATCTA
GTACCACTCATGCCAATACATTTCAAGCAAACGCTTAGTCTGACTAAGAGTATCACTCAAAACCAAATCC
GAAGATTTGAGAGAGAAATGACGCTCAAACAGGCATGCCCTTTGGAATACCAAAGGGCGCAATGTGCGTT
CAAAGATTCGATGATTCACGAAAATCTGCAATTCATATTACTTATCGCATTTCGCTGCGTTCTTCATCGA
TGCGAGAACCAAGAG

10. LD01BII >*Meyerozyma guilliermondii* >FR07618468 (998 bases)

GCAGTACCTCGGTCTAGGCAGGCAGCATCAACGCAGGCTATAACACTTCACCGAAGTAAAGTCACATTCC
TACGCCATTATCCTACCGCCCAAACCGATGCTGGCCCGATAAGCTGCGGGTCACCCCGCCACGAAGGCAA
GGCTCACAAAATATCGAGTCTGATCTCAAACCCTTCCCTTTCAACAATTTCACGTACTTTTTCACTCTCT
TTTCAAAGTTCTTTTCATCTTTCCATCACTGTACTTGTTCGCTATCGGTCTCTCGCCAATATTTAGCTTT
AGATGGAATTTACCACCCACTTAGAGCTGCATTCCCAAACAACTCGACTCTTCGAAAGCACCTTACATAG
AATTGGGCATCTCATCGCACGGGATTCTCACCCTCTGTGACGTCCTGTTCCAAGAAACATAGACAAGAGC
CAACCCCAAGGTTACAATCTTCAAATTACAACTCGGACACCGAAGGCGCCAGATTTCAAATTTGAGCTTT
TGCCGCTTCACTCGCCGCTACTAAGGCAATCCCTGTTGGTTTCTTTTCCTCCGCTTATTGATATGCTTAA
GTTCAGCGGGTATTCCTACCTGATTTGAGGTCAAACTTGTTTGGTTGTTGTAAGGCCGGGCCAACAATAC
CAGAAATATCCCGCCACACCATTCAACGAGTTGGATAAACCTAATACATTGAGAGGTCGACAGCACTATC
CAGTACTACCCATGCCAATACTTTTCAAGCAAACGCCTAGTCCGACTAAGAGTATCACTCAATACCAAAC
CCGGGGGGTTTGAGAGAGAAATGACGCTCAAACAGGCATGCCCTCTGGAATACCAGAGGGCGCAATGTGCG
TTCAAAGATTCGATGATTCACGAAAATCTGCAATTCATATTACTTATCGCATTTCGCTGCGTTCTTCATC
GATGCGAGAACCAAGAGATCCGTTGTTGAAAGTTTTGAAGATTAATTCAAAATTTGACTAACTGTAAAAT
AATTAAATTGGGTTTTGT

11. LD02A >*Wickerhamomyces anomalus* >FR07618498 (602 bases)

GATGGCGGAAGGATCCTAGTCAAAGACGCAGCCCTCGATCCAGACAGGCAATATCAGCAGAAGCTATAAC
ACTCCACCGAAGTGAAGCCACATTCAACTGCCATTATCTTGCCATCCGAATCGATGCTGGCCCAGTGAAA
TACGAGTGCACAACTCAAGAAGAGAAGATAATCGTAAAACACCAAGTCTGATCTAATGCCCTTCCCTTTC
AACAATTTCACGTACTTTTTCACTCTCTTTTCAAAGTTCTTTTCATCTTTCCATCACTGTACTTGTTCGC
TATCGGTCTCTCGCCAATATTTAGCTTTAGATGGAATTTACCACCCACTTAGAGCTGCATTCCCAAACAA
CTCGACTCTTCGATAGCACCTTACATAGGAATGGGCATCTCATCAGACGGGATTCTCACCCTCTATGACG
TCCTGTTCCAAGGAACATAGACAAGAGCCAAACCCAAGGTTACCATCTTCAAATTACAACTCGAACACCG
AAGGTGCTAGATTTCAAATTTGAGCTTTTGCCGCTTCACTCGCCGTTACTGAGGCAATCCCTGTTGGTTT
CTTTTCCTCCGCTTATTGATATGCTTAATTCAGGGGGGGGAA

12. UB04BI >*Rhodotorula mucilaginosa* >FR07618506 (998 bases)

GGAGCCGAATAAAGTTGTTTCAAGGAAGTAGAGCAGGGACATTAGTGAATATAGGACGTCCAACTTAACT
TGGAGTCCGAACTCTCACTTTCTAACCCTGTGCACTTGTTTGGGATAGTAACTCTCGCCCGATAGCGAAC
TCCTATTCACTTATAAACACAAGGTCTATGAATGTATTAAATTTTATAACGAAATAAAACTTTCAACAAC
GGATCTCTTGGCTCTCGCATCGATGAAGAACGCAGCGAAATGCCATAAGTAATGTGAATTGCAGAATTCA
GTGAATCATCGAATCTTTGAACGCACCTTGCGCTCCATGGCATTCCGTGGAGCATGCCTGTTTGAGTGTC
ATGAATACTTCAACCCTCCTCTTTCTTAATGATTGAAGAGGTGTTTGGTTTCTGAGCGCTGCTGGCCTTT
ACGGGCTAGCTCGTTCGTAATGCATTAGCATCCGCAATCGAACTTCGGATTGACTTGGCGTAATAGACTA
TTCGCTGAGGAATTCTAGTCTTCGGACTACAGCCGGGTTGGGTTAAAGGAAGCTTCTAATCAGAATGTCT
ACATTTTAAGATTAGATCTCAAATCAGGTAGGACTACCCGCTGAACTTAAGCATATCAATAAGCGGAGGA
AAAGAAACTATCAGGGATTCCCCTAGTAGCGGCGAGCGAAGCGGGAAGAGCTCAAATTTATAATCTGGCA
CCTTCGGTGTCCGAGTTGTAATCTCTACAAATGTTTCCGCGTTGGACCGCACACTAGGTCTGTGGATACA
GCGCATACTGCTGGAGACCCCCGTACATGCTGCGGACGACCAGCGCTTAGGGATACAGTTTCCAAGATTC
GGCTGTTTAGCAATGCAGCTCAAATTGATGTAGATCATCTGAGCTTAAATATTGGCGGGAATCGATACGT
ACAAGTACCGGTGAGTAAGAGTATAAGACCTTGGCTAAGGAGTTCACAGTACTGAAATCGTGTATGAAGC
TGAAGTCGAACTGCATGC

2.4 Culturing of identified strains into the four stages of the initial Bioprospecting steps

To determine whether the identified strains could survive the designed screening consisting of the four stages of different culture media, we will again run all of the strains onto the vigorous steps. Each strain were cultured using the exact method, each stage independently without the have them following the sequence.

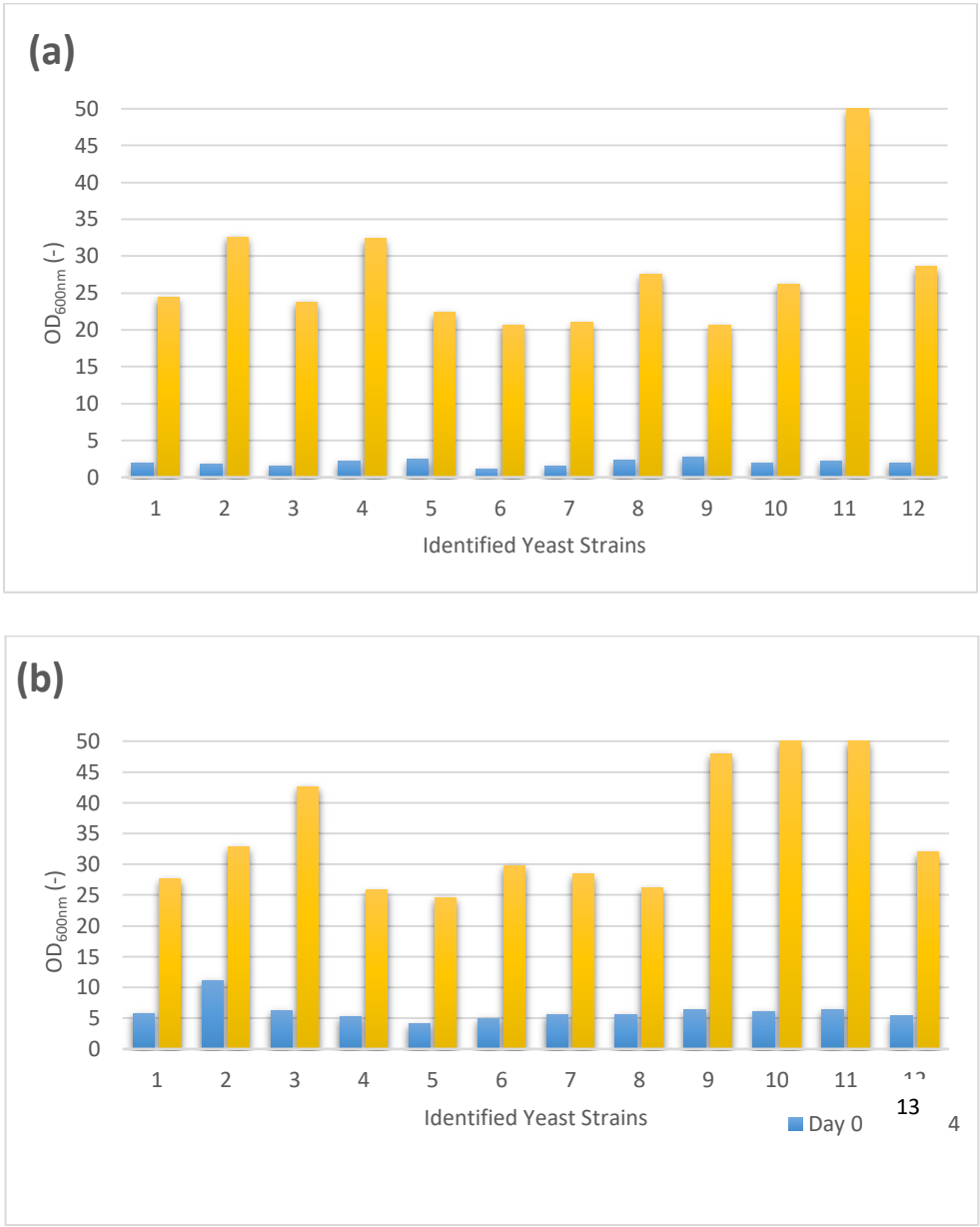


Figure 2.11: The 12 strains of yeasts cultured in (a) MML_A (b) MML_B

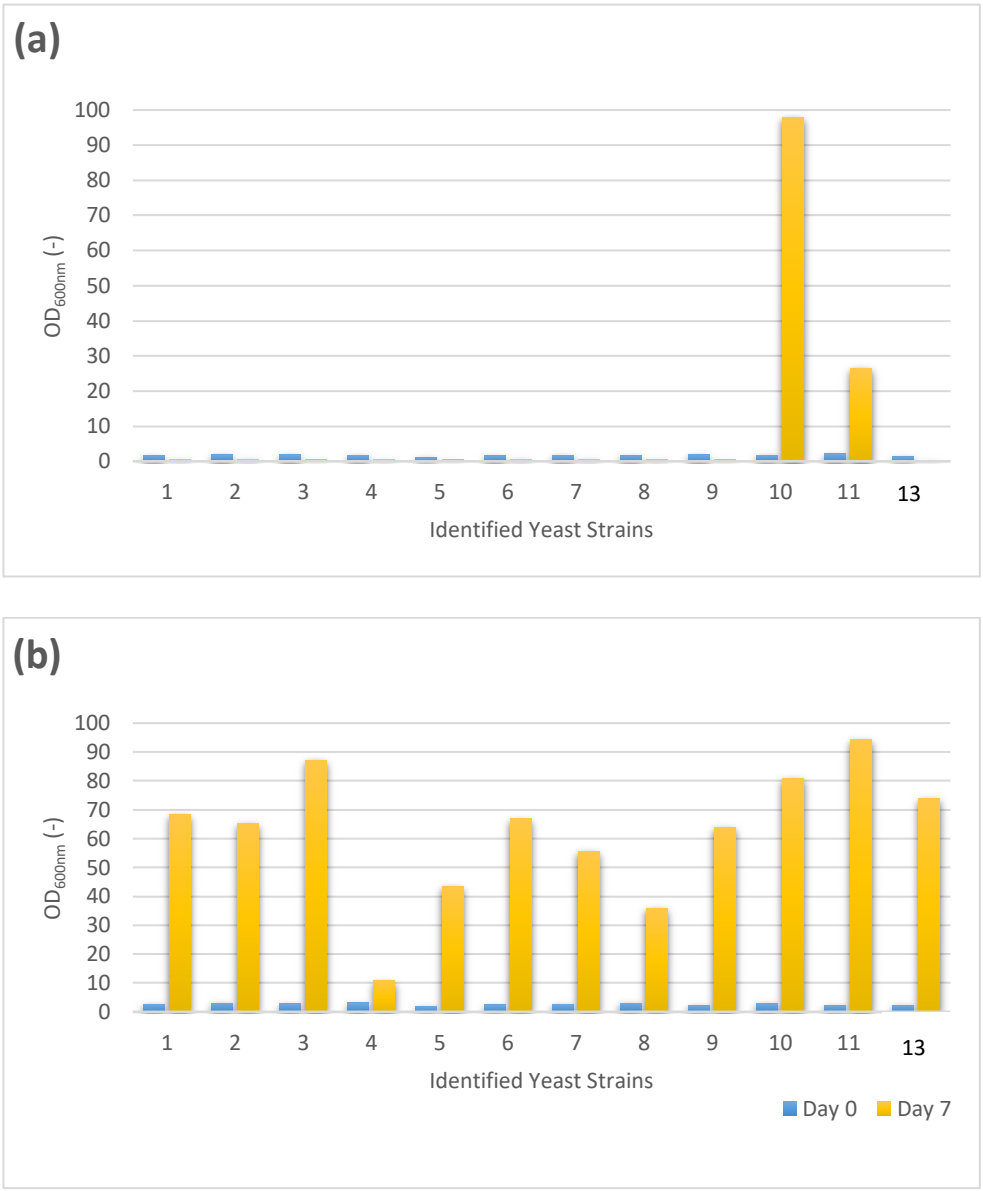


Figure 2.12: The 12 strains of yeasts cultured in (a) MMI_A (b) MMI_B

In the first set of yeast strains cultured in the MML_A, they all grew 20-fold or more from day 0 [Fig. 2.11(a)].The high glucose content would give the best resource for such growth. Strain 11, which is not from the *Metschnikowia* group grew in the highest rate. While in the basic condition, MML_B, they only showed 5-10 fold of growth rate, strains 3, 9, 10 & 11 showing fastest growth where only strain 3 is from the *Metschnikowia* group [Fig. 2.11(b)]. . The pH 9.0 negates all of the acidic inhibitors destructive properties.

In MMI_A, the inhibitors exhibit its destructive capabilities and killed all strains except 10 & 11, both not of the *Metschnikowia* group. While in the basic culture, desirable growth shown in 7 days, except for strain 4. Again strains 3 & 11 showed the best growth rate.

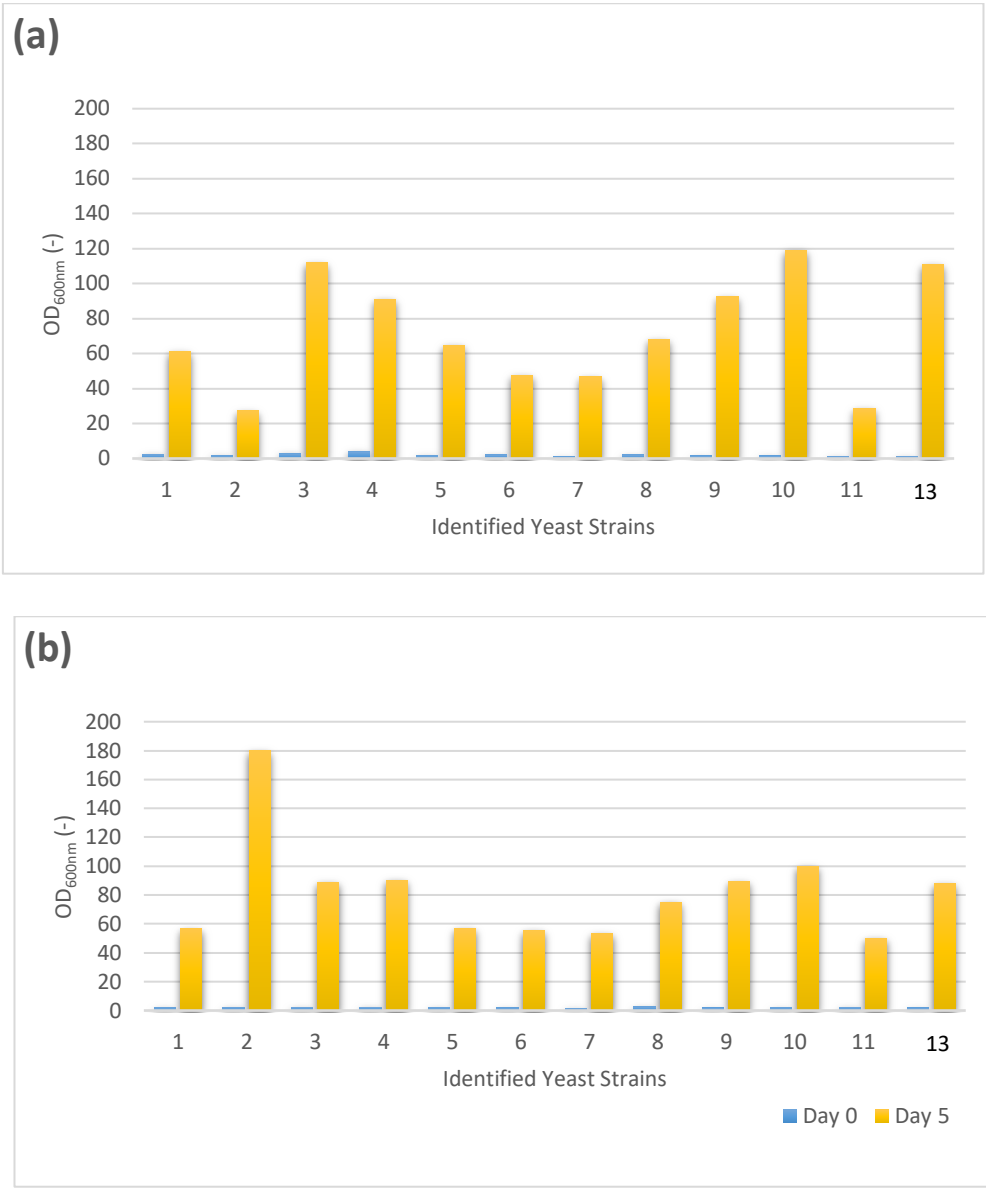


Figure 2.13: The 12 strains of yeasts cultured in (a) MMX_A (b) MMX_B

The growth of all 12 strains are positive in both MMX conditions. In acidic condition, all grew 10-fold except strains 2 & 11 [Fig. 2.13(a)]. Strains 3, 10 & 13 showed greatest growth in MMX_A, while strain 2 in MMX_B. Strain 2 has the highest growth in MMX_B, but did not do well in acidic condition. This strain also showed its better survival in basic condition rather than acidic conditions in MMI_B Strain 11 showed very slow growth after 5 days in MMX [Fig. 2.13(b)], as oppose to its excellent growth in all the other cultures in both acid & basic conditions. This show that the strain have lower capabilities in utilizing xylose.

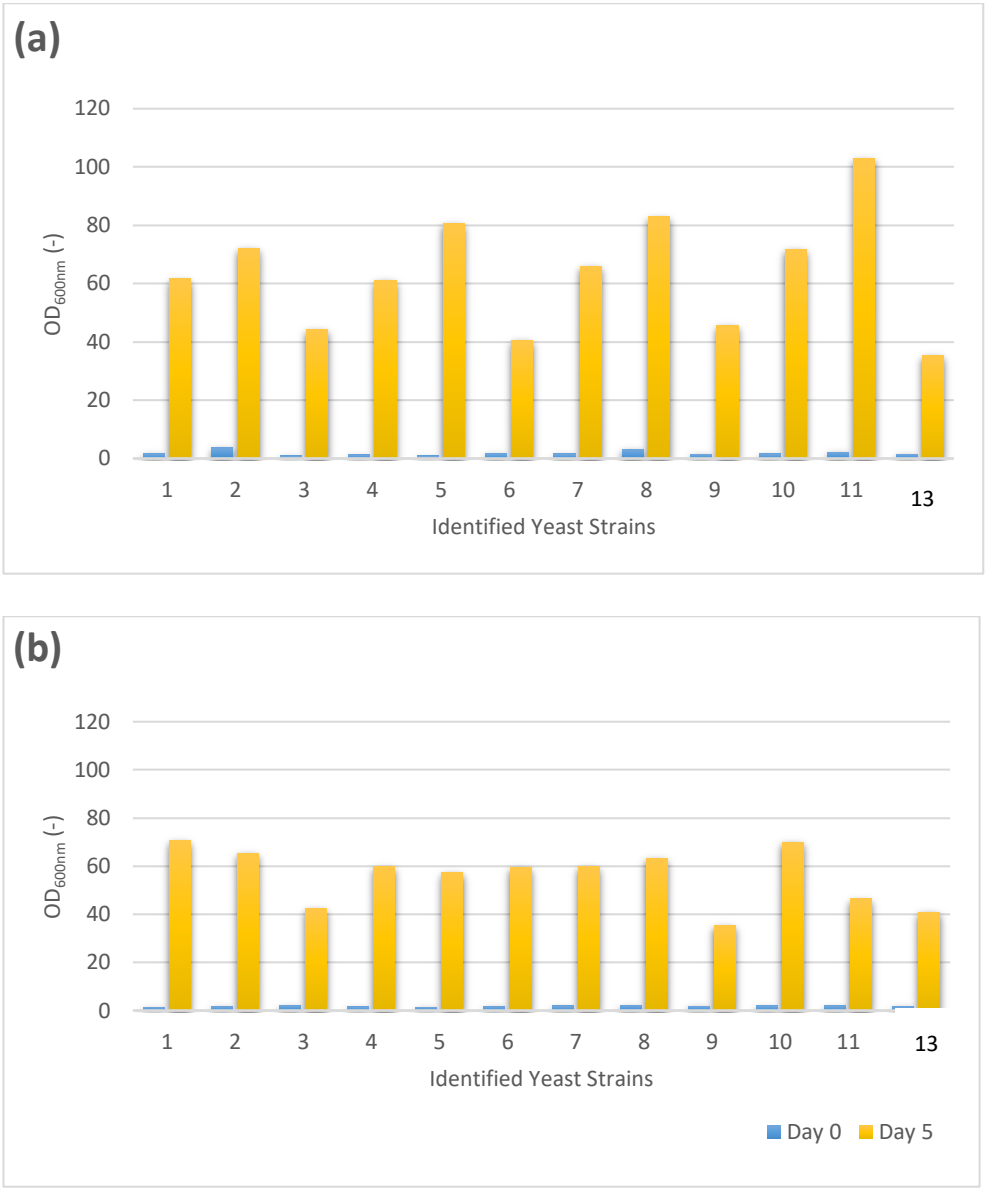


Figure 2.14: The 12 strains of yeasts cultured in (a) $MMAC_A$ (b) $MMAC_B$

In the last set of culture MMAC, in acidic condition, strain 11 grew 50-fold efficiently utilizing either arabinose or cellobiose or both. Arabinose is an aldopentose, which most yeast could not hydrolyse [116]. Cellobiose on the other hand, is a disaccharide consisting of two β-glucose. Therefore, the ability for the yeasts strains to hydrolyse the disaccharide is crucial. In earlier studies showed tha the hydrolyzation of cellobioase takes more than 5 days in yeasts, such as *Hansenula polymorpha* & *Pichia stipitis* [117]. So there is least of possibilities that the yeasts in these experiments are able to break down cellobiose in shorter time. Further studies on the ability of these yeast to consume these less common sugars.

2.5 Conclusions

A novel bioprospecting approach was used to determine whether yeasts local to the area could be found that are suitable for the industrial production of bioproducts from lignocellulosic hydrolysates. The yeasts were screened applying a staggered culturing method, filtering them through harsh conditions found in lignocellulosic hydrolysates. While there were presumably hundreds of strains living in the environments tested, only the most resilient were able to survive this screening.

Altogether 24 strains were identified from PCR-DNA sequencing method. Some overlaps of exact strains were observed. As such 12 strains were identified. As in many DNA-sequenced Bioprospecting outcomes, new unidentified species were anticipated. Three different strains of *Metschnikowia aff. chrysoperlae* were unearthed at different locations of the University. What is more exciting is that these strains a genetically closely related species to *M. pulcherrima*, a promising organism for oil production. Though it remains to be seen whether the rest of the species in this collection are oleaginous.

While there are indications that the 12 yeast are suitable candidates for biotechnology, the optimal conditions for growth and oil production must be found. In the next chapter, the strains will be grown in different feedstock environments that simulate different waste lignocellulose compositions, including inhibitors.

Chapter 3

Assessment of the most promising yeast strains for lipid production through culturing on model lignocellulosic feedstocks

3.1 Introduction

In the previous chapter, a number of yeast strains were isolated that were resilient to inhibitors and could survive being cultured at low pH. In addition, these strains were proven to have the ability to grow on less common sugars, such as xylose, and in low nitrogen conditions [118]. These significant attributes would be a huge benefit in culturing on an industrial scale, and could even lead to high lipid production under stressful conditions.

While yeasts flourish when cultured on glucose or sucrose, this is generally not cost effective in most parts of the world due to a lack of sugar-rich crops. One solution to provide for the industrial production of microbial lipid would be to use agricultural residues as the carbon source, where millions of tonnes are produced yearly in Europe and the U.S. [39]. These lignocellulose waste materials are widely available throughout the world where there is a traditional agricultural base [119]. The most abundant feedstock worldwide is rice straw, followed by wheat straw and sugarcane bagasse, as illustrated in table 3.1 [37].

Agricultural wastes	Availability (million tons)	Estimated bioethanol potential (gl)
Wheat straw	354.34	104
Rice straw	731.3	205
Corn straw	128.02	58.6
Sugarcane bagasse	180.73	51.3

Table 3.1: An estimate on the global lignocellulose availability. Adapted from Sarkar et al. [4]

The European Union (EU) generated 139 million tonnes of crop residues in 2012, mainly from wheat production [120]. Wheat yields 28% of the total harvested agricultural area in EU, and wheat straw production is estimated to be 156 million metric tonnes per year worldwide [121].

3.1.1 Lignocellulose depolymerization

Generally, lignocellulose is composed of cellulose (30-50%), hemicellulose (15-35%) and lignin (10-20%). Cellulose and hemicellulose, representing roughly 75% of the total, are interwoven in covalent and hydrogen bonds to produce a highly stable structure that is insusceptible to most degradation treatments [54].

Cellulose is an unbranched homopolysaccharide consisting of D-glucopyranosyl units. Hemicelluloses are branched heteropolysaccharides which consist of hexose and pentose sugars. Lignin, comprises of phenylpropane units linked together by a variety of types of linkages [62].

Arguably, the greatest economic barrier and challenge currently is to cost-effectively breakdown the lignocellulose suitable for microbial growth [122]. There are two main methods to convert lignocellulose into biofuels and bio-products: thermochemical and biochemical [33]. The thermochemical conversion process is shorter but requires more energy input, while the biochemical conversion, in theory, has higher returns with low energy consumption [52].

For example the conversion of lignocellulosic biomass into bioethanol consists of the mechanical and chemical pretreatments; hydrolysis of the solubilised hemicellulose and cellulose by enzymes; fermentation of the saccharides; and distillation to purify the bioethethanol [53]. The original mechanical and chemical pretreatment stages of the lignocellulose are basically to break compact structures so that substrates can penetrate, therefore maximizing the amount of cellulases that come in contact with the cellulose [54] (Fig. 3.1). Pretreatment of wood-derived lignocellulosic is essential for ensuing an effective fermentation to ethanol, as wood tends to be more stubborn than normal crop plants because of the higher lignin content, the intricate ultrastructure and the increased difficulty in hydrolysis of the structural polysaccharides [55]. The pretreatment process “breaks up” the structural walls [56].

The physical pretreatment stage includes milling, extrusion or microwave treatments. They are usually combined to produce the best results. Milling, though will produce optimum chip sizes, is the most inefficient in energy usage. While, extrusion, a thermo-physical treatment, is costly but does not generate substantial levels of inhibitors. Microwave technology has a short processing time while using less energy [57].

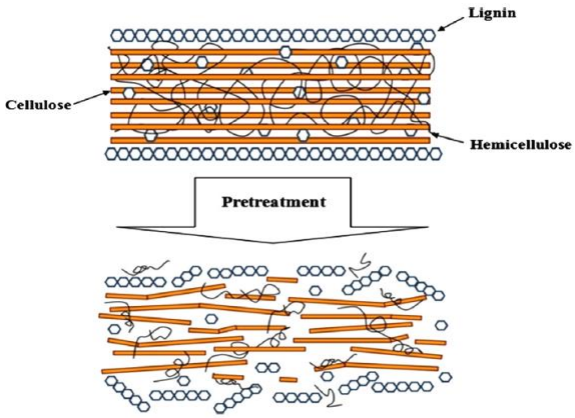


Figure 3.1: Schematic pretreatment of lignnocellulosic material taken from Mood, S.H., et al., 2013 [123]

After pretreatment, various cellulases can be used to produce the monosaccharide and disaccaharides that can be readily fermented to fuel molecules. Cellulase enzymes carry out the enzymatic hydrolysis and are highly specific in production of the reducing sugars including glucose [58]. The main sugars produced from this process are glucose, xylose, cellobiose and arabinose. All of which must be fermentable by the organism of choice, to produce bioethanol.

The processing aspect to produce lipid using microbes would be identical up to the yeast fermentation. In this chapter, four common lignocellulosic sources were selected; wheat straw, corn Stover, sugarcane bagasse and palm kernel cake, as the basis of the model compounds to culture the selected yeasts on.

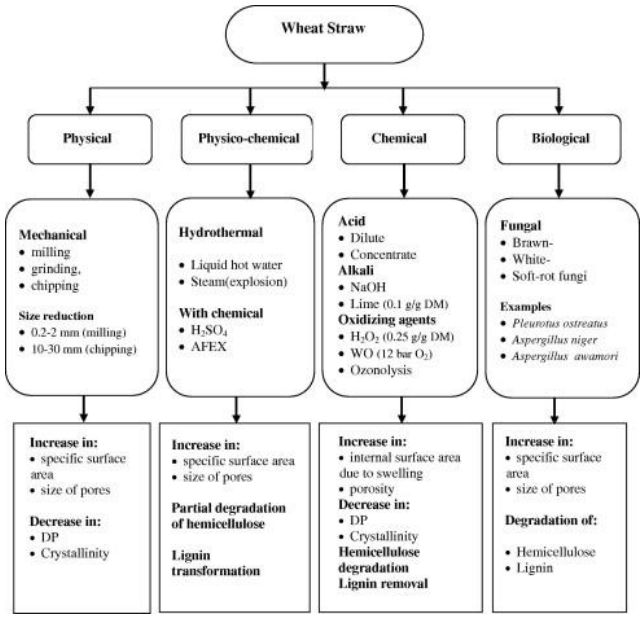


Figure 3.2: The most common pretreatment methods used on wheat straw and their possible effects, adapted from Talebnia (2010)

3.1.2 Lignocellulosic Inhibitors

As lignocellulose goes through the aggregated pretreatments, the sugars produced can also break down further into inhibitory by-products (Fig. 3.3). The pretreatment usually break down the hemicellulose into pentose and hexose sugars, these can then further break down into sugar acids, acetic acid, formic acid, levulinic acid, furan aldehydes, 5-hydroxymethylfurfural (HMF) and furfural [62]. After hydrolysis of lignocellulose polysaccharides, lignin generally remains intact as a solid, though a small amount is degraded to phenolics and other aromatic compounds under especially harsh conditions [62].

The inhibitory compounds are divided into three categories based on their source as above. They are weak acids, furan derivatives, and phenolic compounds [64]. All the inhibitors can hinder the biological processes in different parts of the process [65-67].

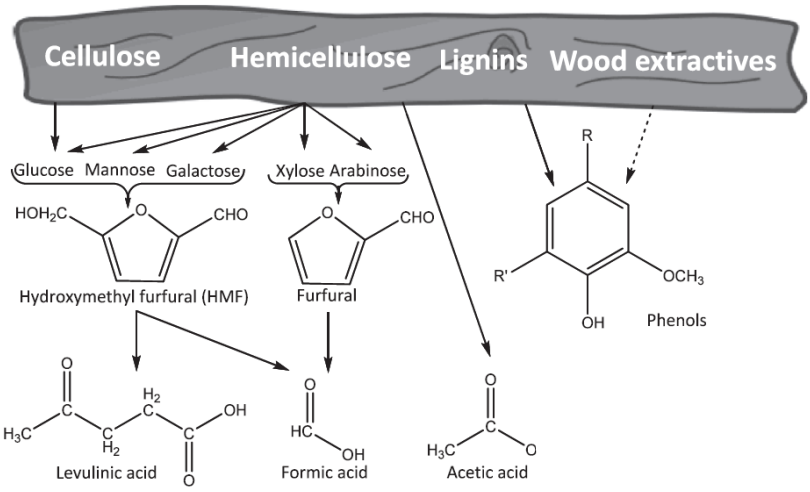


Figure 3.3: Formation of inhibitors from Lignocellulose, taken from reference [62]

3.2 Aims and Objectives

The aim of the first part of this chapter was to grow all (12) surviving strains from the staggered screening methods of in chapter 2 on simulated conditions of different hydrolysates with different degrees of inhibitors and pHs. The simulated hydrolysates, namely: wheat straw; corn Stover; sugarcane Bagasse; and palm kernel cake. The four different degrees of inhibitors entail no-inhibitors, low-inhibitors, medium-inhibitors, and high-inhibitors. The surviving strains have been proven to be resilient to high, low pH and also medium inhibitors concoctions. Therefore, we would like to further investigate the range of conditions mentioned that they best require.

We would like to distinguish one or two strains that would fit all resilient traits that deemed the strain to be compatible with the harsh conditions of the depolymerisation of lignocellulose along with all its distinct carbon content and inhibitors.

In the second part of the chapter, the aim is to determine the oleaginous potential and suitability of the lipid produced as a palm oil substitute, for each strain, quantified over two inhibitor levels and pHs. We would also want to select a strain that can produce lipid content closest to palm oil.

3.3 Results and Discussion

3.3.1 Culture conditions

In chapter 2 the range of yeasts were selected based on inhibitor tolerance, tolerance to lower temperatures and for being able to metabolise a wide range of sugars. To assess the strains for lipid production, four different lignocellulosic hydrolysates were modelled, through combining the pure monosaccharide sugars in different amounts. The lignocellusoic residues were selected due to the abundance across the world, the residues and conditions used were therefore:

1) Typical hydrolysates from four feedstocks from across the globe: wheat straw; corn Stover; sugarcane bagasse and palm kernel cake (as shown in table 3.3).
2) Temperatures of 20 °C and 25 °C
3) Different levels of inhibitors. Four degrees (no-inhibitors, low-inhibitors, medium-inhibitors, high-inhibitors) of the five more apparent inhibitors: Furfural, 5-(Hydroxymethyl) furfural (5-HMF), Acetic Acid, Formic Acid and Levullinic Acid.
4) 9 different pH's: pH 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 7.0.

The 12 strains isolated in chapter 2 were then cultured in 96 well plates, at 180 rpm, over 5 days under N limiting condition (table 3.2).

Table 3.2: The optimization specification matrix for 12 strains.

Temperature	**20 ° C / 25 °C**															
Sugars	Wheat Straw				Corn Stover				Sugarcane Bagasse				Palm Kernel Cake			
Inhibitors level	None	Low	Medium	High	None	Low	Medium	High	None	Low	Medium	High	None	Low	Medium	High
pH	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5	1.5
	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0	2.0
	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5	3.5
	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0	6.0
	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0	7.0

Table 3.3 Model hydrolysate composition used in this study.

Hydrolysate	Sugar Content	References
Wheat Straw	**Xylose, 12.80 g/L** **Glucose, 1.70 g/L** **Arabinose, 2.60 g/L**	**[34]**
Corn Stover	**Xylose, 9.09 g/L** **Glucose, 2.13 g/L** **Arabinose, 1.01 g/L**	**[124]**
Sugarcane Bagasse	**Glucose 13.92 g/L** **Xylose 7.123 g/L** **Arabinose 0.647 g/L** **Glucuronic acid 0.6414 g/L**	**[125]**
Palm Kernel Cake	**Glucose 2.31 g/L** **Xylose 0.78 g/L** **Arabinose 0.33 g/L** **Galactose 0.57 g/L** **Mannose 10.71 g/L**	**[126]**

3.3.2 Model lignocellulose hydrolysates

The growth of the 12 yeast strains on the wheat straw model hydrolysate (12.8 g/L xylose, 1.7 g/L glucose, 2.6 g/L arabinose) at both 20 °C and 25 °C is shown in figure 3.4 and figure 3.5. The second hydrolysate model, corn stover (9.09 g/L xylose, 2.13 g/L glucose, 1.01 g/L arabinose) growth at both 20 °C and 25 °C is presented in figure 3.6 and figure 3.7. The third formulated in this experiment, sugarcane bagasse (13.92 g/L glucose, 7.12 g/L xylose, 0.65 g/L arabinose, 0.64 g/L glucuronic acid) growth at both 20 °C and 25 °C is displayed in figure 3.8 and figure 3.9. Lastly, figure 3.10 and figure 3.11 represented the growth of the 12 strains at 20 °C and 25 °C on the Palm Kernel Cake model (2.31 g/L glucose, 0.78 g/L xylose, 0.33 g/L arabinose, 0.57 g/L galactose, 10.71 g/L mannose)

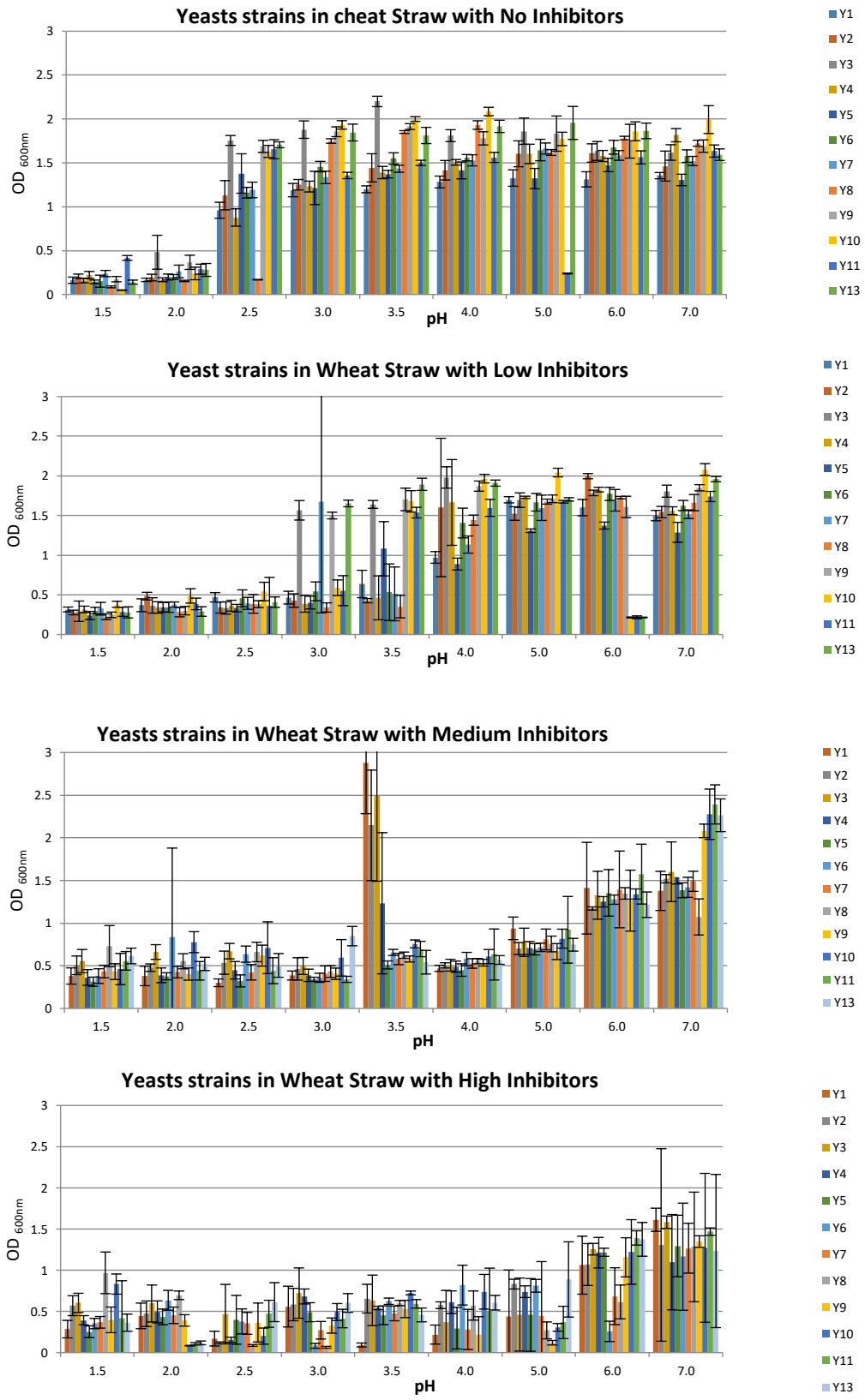


Figure 3.4: The 12 strains of yeasts cultured in wheat straw sugar content simulation; with four different inhibitors levels; in nine pH's; and at 20°C temperature.

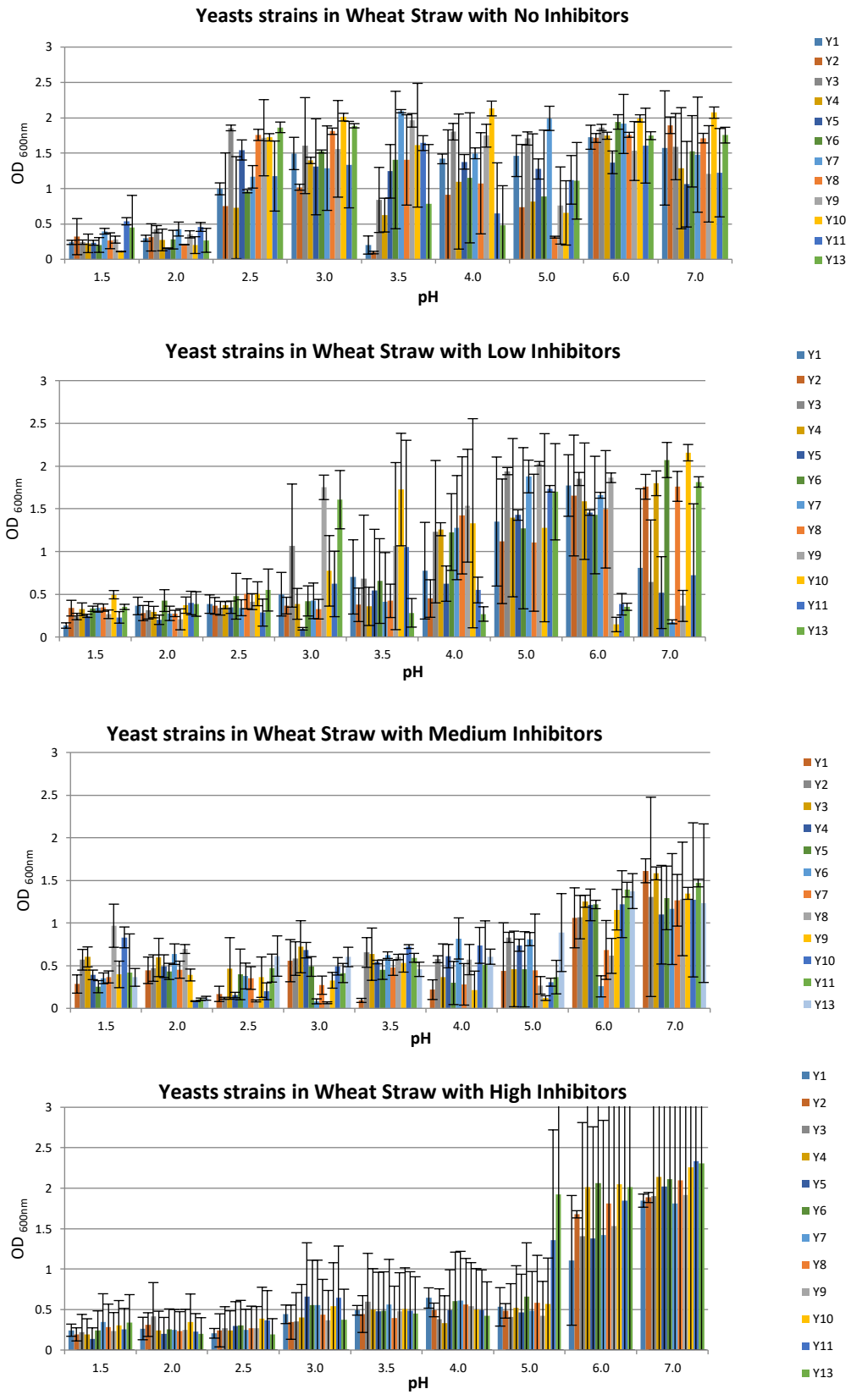


Figure 3.5: The 12 strains of yeasts cultured in wheat straw sugar content simulation; with four different inhibitors levels; in nine pH's; and at 25°C temperature.

All the yeast species behaved in a similar manner, as might be expected from being isolated from the same bioprospecting regime. At 20 °C, with no inhibitors reasonable growth for all the species was observed for all species down to pH 2.5-3.0. The relatively low amount of glucose did not seem to effect the yeast, with all species thriving on the xylose / arabinose rich media. However, interestingly, just because the species can survive at low pH, under increasing inhibitor concentrations the pH that gives reasonable growth increases. This summarises as the higher the inhibitors concentrations, the higher the pH must be for the strains to give better growth rate. For example, with a low level of inhibitors this is 3.5-4.0, with medium or high levels of inhibitors then reasonable growth is only observed at pH 6.5-7.0. The same trend is seen at 25 °C, though the data is subject to higher errors, presumably as the higher temperature gives more evaporation and condensation. This would decreases the water content in a few of the technical repeats and gives a big difference between readings.

All the inhibitors can hinder the biological processes in different parts of the process. For example, undissociated weak acids are lipo soluble and dispersed across the plasma membrane, the low dissociation of H^+ into the cytosol inhibits growth of yeasts and bacteria [65]. The cell reproduction also decreases as the pH decreases [66]. These acids have been shown to severely reduce ethanol production in *S. cerevisiae* [67].

One theory on the negative impact of acids on cellular growth is that the ATP hydrolysis is at a high, hence the proton-pumping capacity is exhausted by the cell (to maintain the intracellular pH) [66]. Alternatively, the anion accumulation theory has also been put forward. In this theory high anion accumulation in the cell will create internal acidification and directly interfere with the cell, inhibiting growth [68].

This increasing intolerance to inhibitors at low pH reinforces the idea that there is a co-effect between the inhibitors and that a blend of inhibitors including acids is worse for growth than individual components. The model corn stover media is a similar concentration though with an increased glucose to xylose ratio. All the yeasts were again screened in 96 well plates, at a range of pHs and 20 and 25 using this media (Fig 3.6 and fig. 3.7).

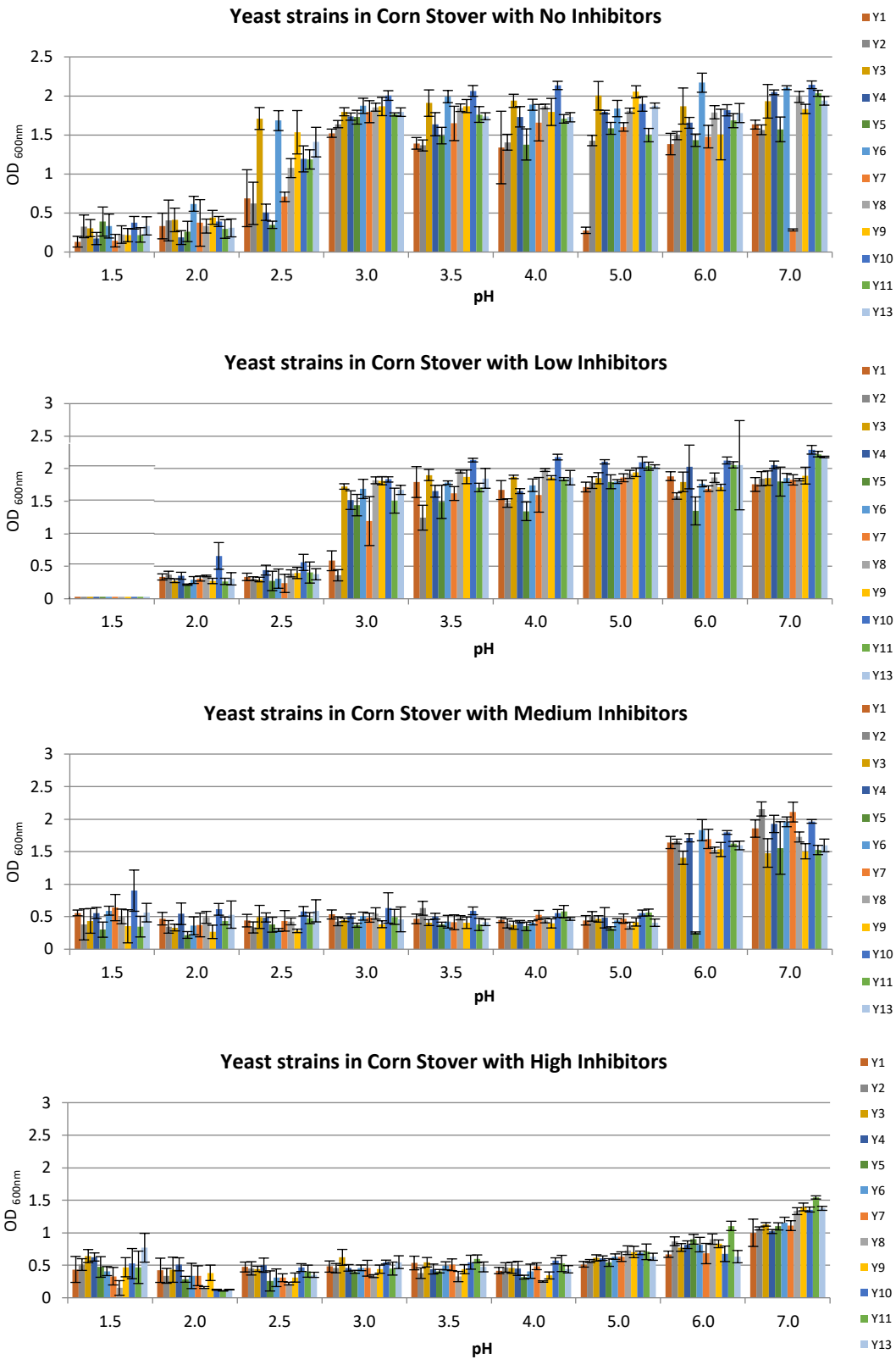


Figure 3.6: The 12 strains of yeasts cultured in corn Stover sugar content simulation; with four different inhibitors levels; in nine pH's; and at 20°C.

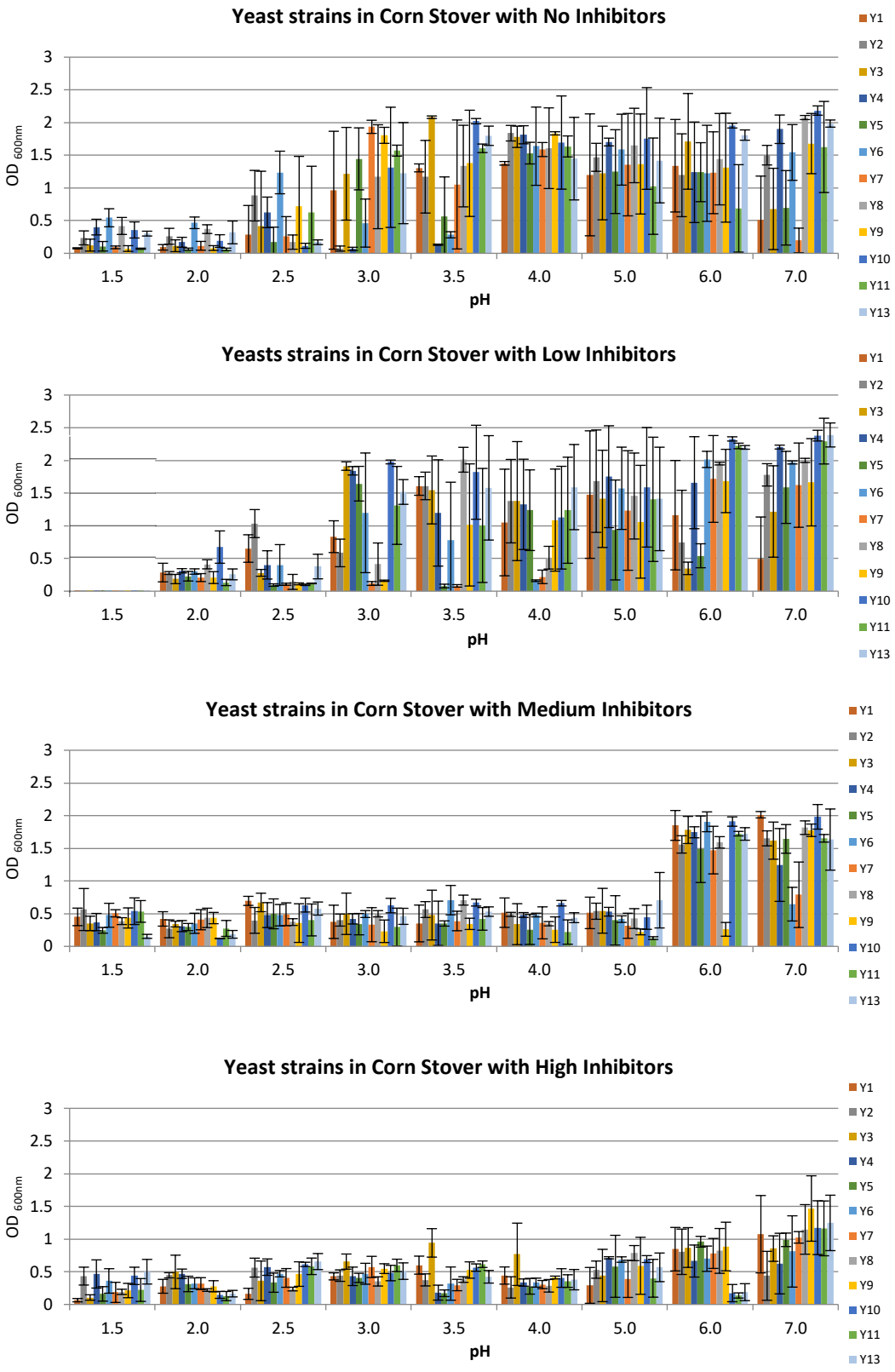


Figure 3.7: The 12 strains of yeasts cultured in Corn Stover sugar content simulation; with four different inhibitors levels; in nine pH's; and at 25°C.

Similar results were observed in the model corn Stover hydrolysate at 20˚ with the model wheat straw, where limited growth is observed for all of the species at low pH of 1.5 – 4.0. When a medium level of inhibitors are present in the medium then only pH 6.5 or 7 supports reasonable growth. This pattern is repeated for the cultures grown at 25°C. At both temperatures growth was severely reduced for all species, over all pH ranges at high inhibitor loadings.

In the corn Stover simulated condition and low inhibitors, the pH 1.5 set of experiments had been entirely unsuccessful in two trials, we assume that there is some error in the concoction of media. At both occurrence, they showed very high growth rate.

Both of these models have predominantly xylose in the relatively low sugar concentrations. However, the model hydrolysate for sugarcane bagasse contains over 60g/L of sugar, with an approximate ratio of 2:1 glucose to xylose. All 12 species were cultured on the model media at 20 and 25 °C, with a variety of inhibitor loadings (fig. 3.8 and fig. 3.9). The also show the similar trends of better surviving in higher pH and lower inhibitors concentration.

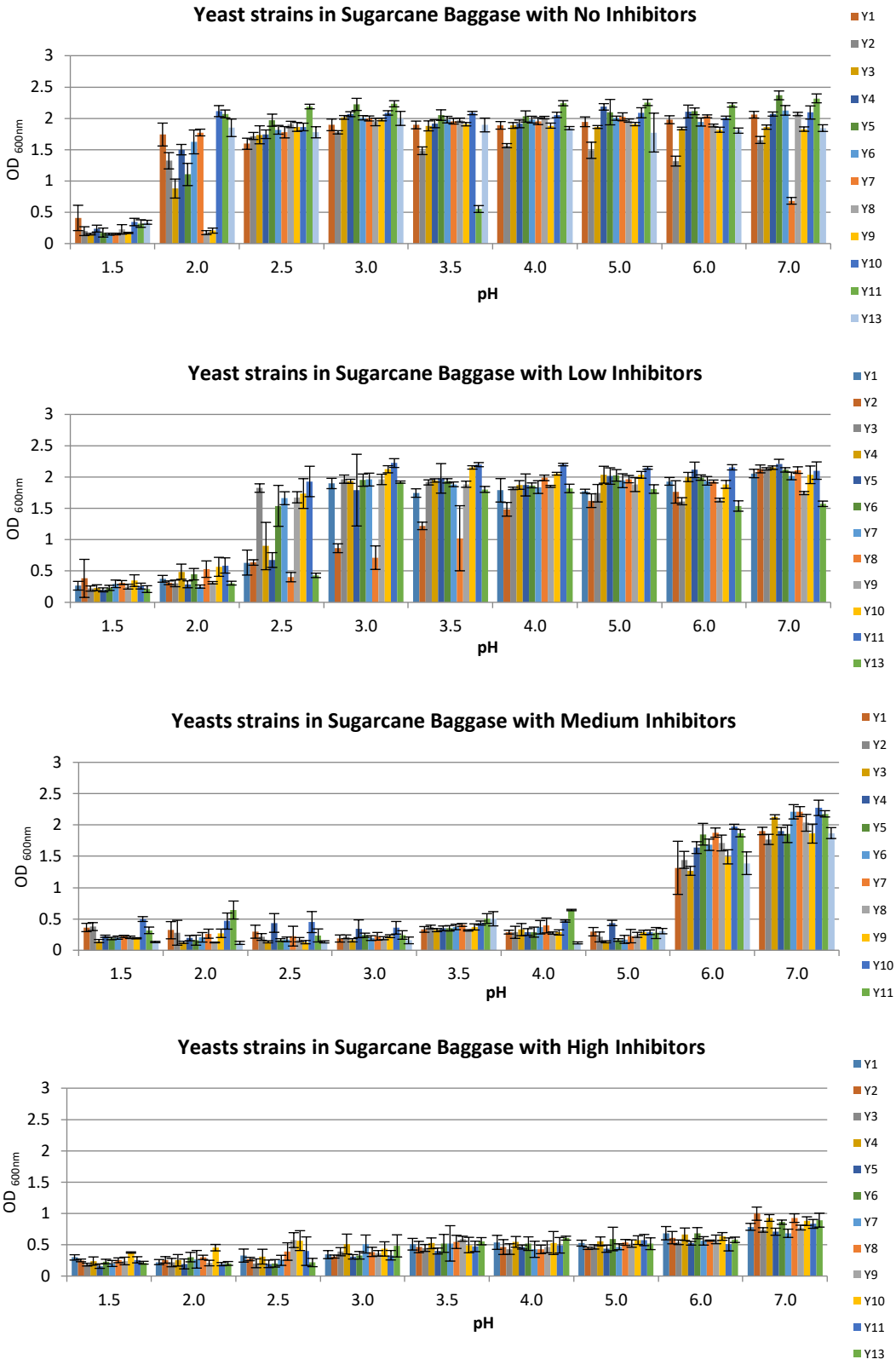


Figure 3.8: The 12 strains of yeasts cultured in Sugarcane Bagasse sugar content simulation; with four different inhibitors levels; in nine pH's; and at 20°C.

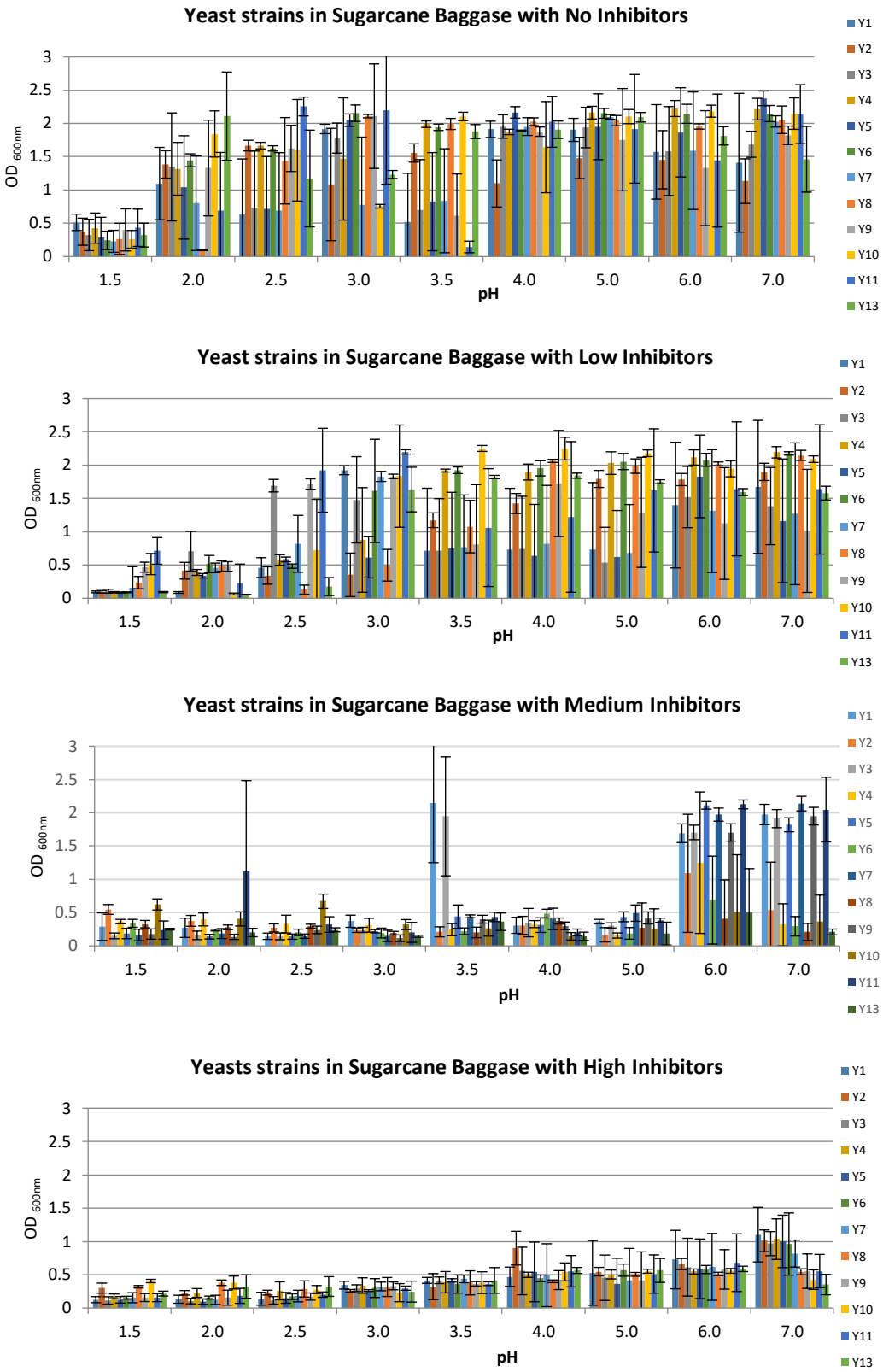


Figure 3.9: The 12 strains of yeasts cultured in Sugarcane Bagasse sugar content simulation; with four different inhibitors levels; in nine pH's; and at 25° C.

At 20 °C, with the higher sugar and glucose loading the growth was substantial for all the species tested, with the exception of strain 13, as low as pH 2. Even at low inhibitor concentrations strong growth is observed at the lower pHs, unlike when culturing in the medias with lower sugars. This effect was similar for the cultures at 25 °C. This demonstrates that there is a complex pay off between the richness of the media in terms of sugar concentration, the inhibitor concentration and the pH. The sugar concentration is reliant on the feedstock but also on the chemical and biological processing, with longer more expensive treatments yielding higher sugar concentrations. However, this is offset by the amount of inhibitors that are formed in these lengthier production processes.

The reason to culture the yeast at a low pH, is to avoid contamination of the yeast through invasive species. However, despite these yeasts being selected to be able to cope at low pH and with high inhibitors, there is a severe trade-off between pH and inhibitors concentration factors, even at high sugar loadings. This may not have too large an impact on the process however, with high inhibitor levels in the original hydrolysate probably having a similar effect to the low pH in keeping the culture sterile. Similarly, if the hydrolysate coming into the process has low inhibitors then the culture could be run at a lower pH to ward off invasion. The industrial production of lipids from these types of yeasts is therefore likely to be a responsive, iterative process.

Finally, the model palm kernel cake hydrolysate was assessed (figure 3.10 and 3.11). This media contained a wider range of sugars, with neither glucose nor xylose being present in large concentrations, the predominant sugar was instead mannose. Again, the pattern of survival maintains the same as the other three prior experiments. At 20˚C, very clean consistency of such pattern. However, at pH3.0, 3.5 & 4.0 in low inhibitors medium, yeast strains 9,10,11,13 showed very similar growth rather than having a big diferrence of growth between the two pHs as shown by the other strains which are all in the *Metschnikowia* family. We can say that the 4 strains can tolerate better in lower pH and low inhibitors rather than the others mentioned. A very low growth was detected in strain 8 at pH 5.0 in low inhibitors, the same was observed in medium inhibitors at pH 6.0 for strain 1 and at pH 7.0 for strain 6. These low growth is incongruent with the pattern of higher survival at lower inhibitors & higher pH. We assumed that there are some error in loading of the yeasts during transfer (human error).

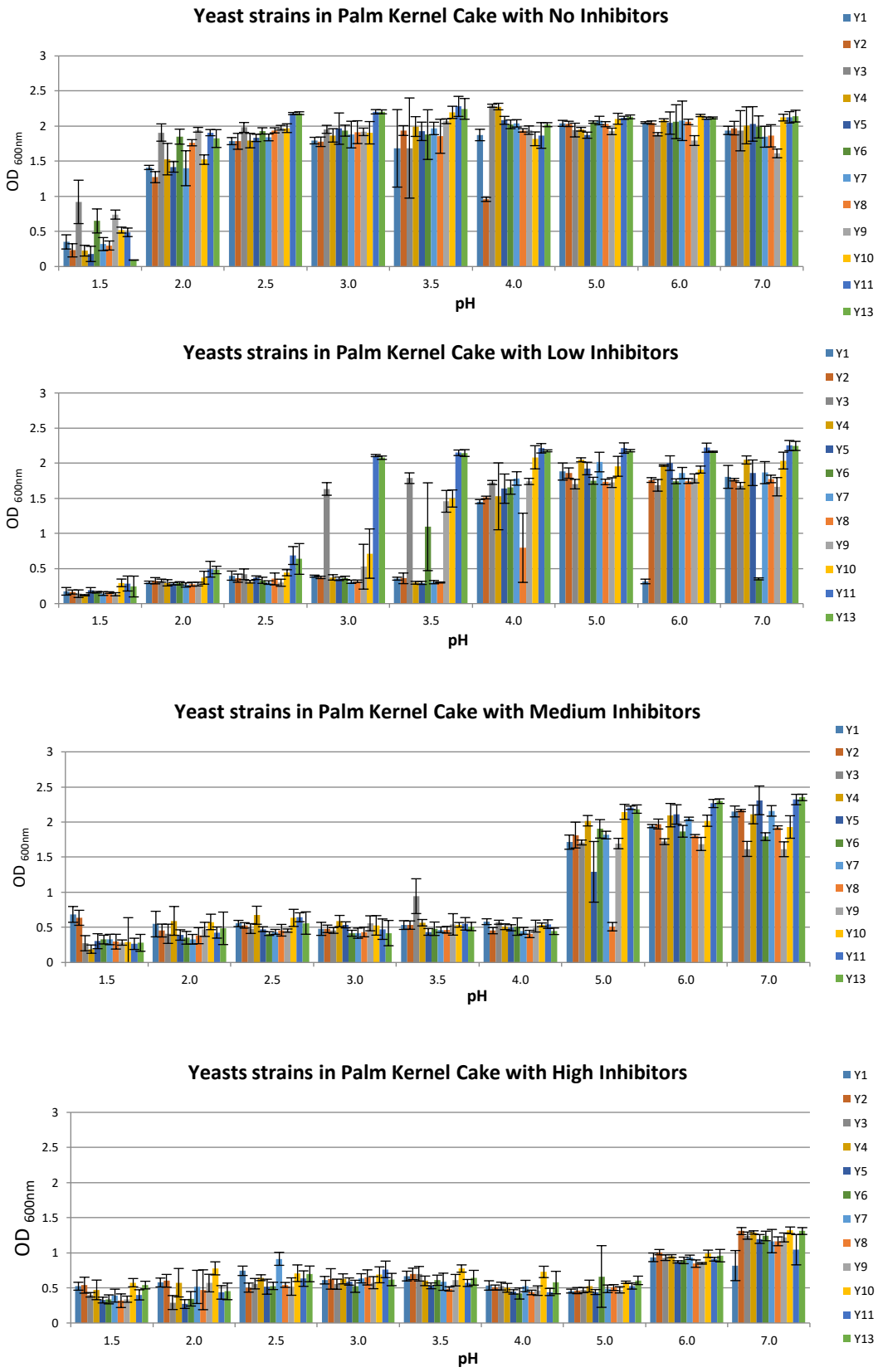


Figure 3.10: The 12 strains of yeasts cultured in Palm Kernel Cake sugar content simulation; with four different inhibitors levels; in nine pH's; and at 20 °C.

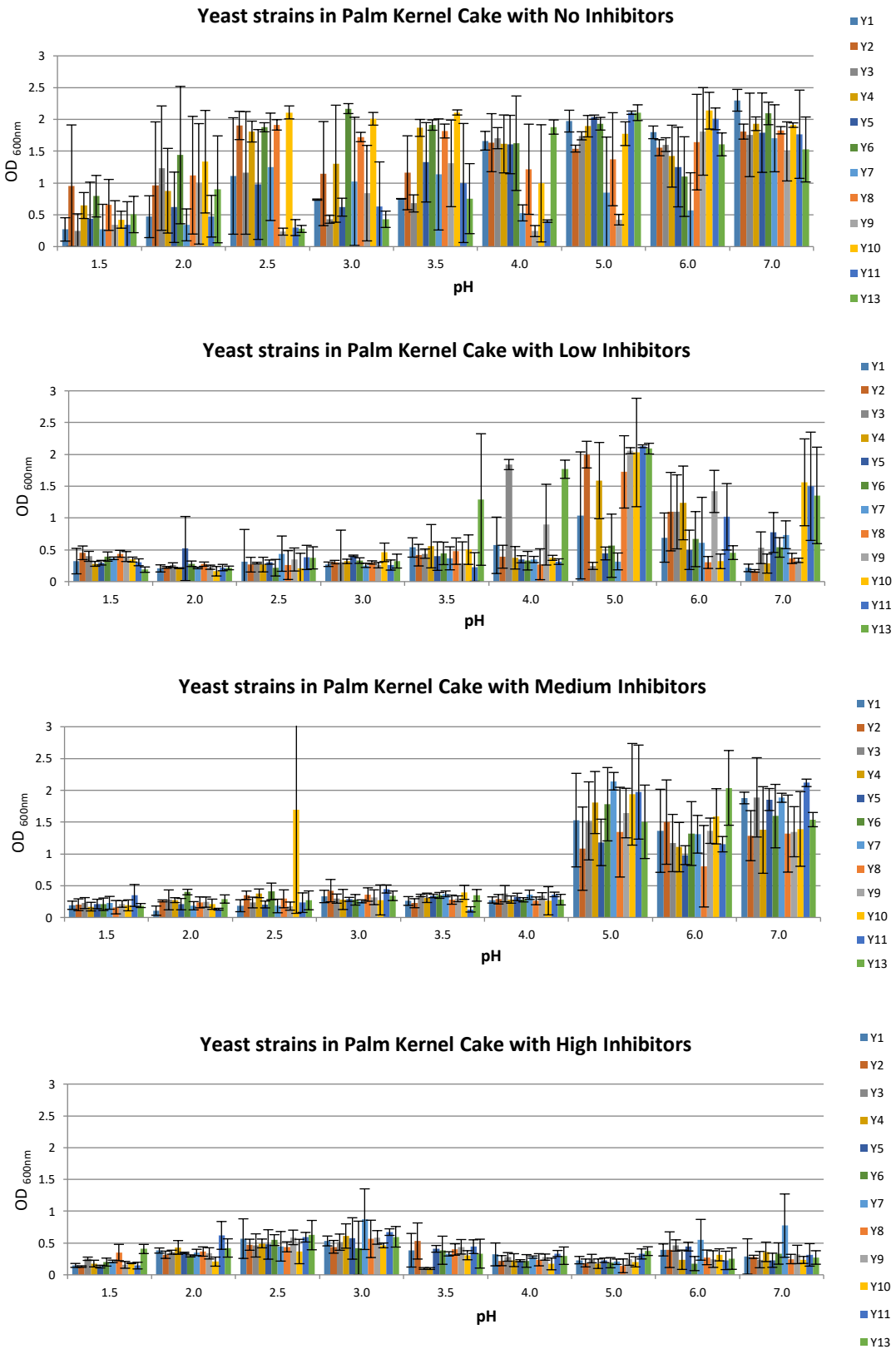


Figure 3.11: The 12 strains of yeasts cultured in Palm Kernel Cake sugar content simulation; with four different inhibitors levels; in nine pH's; and at 25°C.

Similarly, to the high sugar media, the yeasts grew well at all pH ranges except pH 1.5 at 20 °C when no inhibitors were present. This demonstrates that the yeasts could all metabolise mannitol, as well as the glucose and xylose that the original bioprospecting used. Similarly to the other cultures however, an increase in inhibitors increased the lowest possible pH with high inhibitors demonstrating very little growth. The results were similar for the cultures at 25 °C.

Interestingly, there was little difference in the growth across species and rather general rules can be extrapolated for all the strains, in terms of growth, these are:

1. The biggest trade off is between the inhibitor levels and pH, where high inhibitors reduce the pH range that these yeast thrive at.
2. High sugar concentrations mitigate this effect, but not to a large degree, and rather it is more important to match the correct pH to inhibitor level.
3. Mannitol can be used effectively despite not being present in the original bioprospecting. This is evident in the PKC sugar model experiment.

As all yeasts showed similar growth under these model hydrolysates, all the yeasts were taken on to determine if any had oleaginous characteristics.

3.3.3 Oleaginous behaviour

One of the key research aims was to determine if any of the yeasts that showed potential to be cultured under industrial biotechnology conditions could produce a saturated lipid, suitable as a palm oil substitute. To be suitable the yeasts must be able to produce over 20% of their dry weight in lipid.

To assess this the yeasts were cultured in a model media with a glucose to xylose ratio of 2:1. Based on the findings from the previous section two media concentrations were selected. Culturing with low and medium inhibitors, at both pH5 and pH6. All yeasts were cultured at the 2 different temperatures, 20 °C & 25 °C (fig 3.12). The cultures were stopped after 5 days, as this is seen as the longest duration for a lipid production process, with any time beyond this, would incur higher costs in running the fermenters [127].

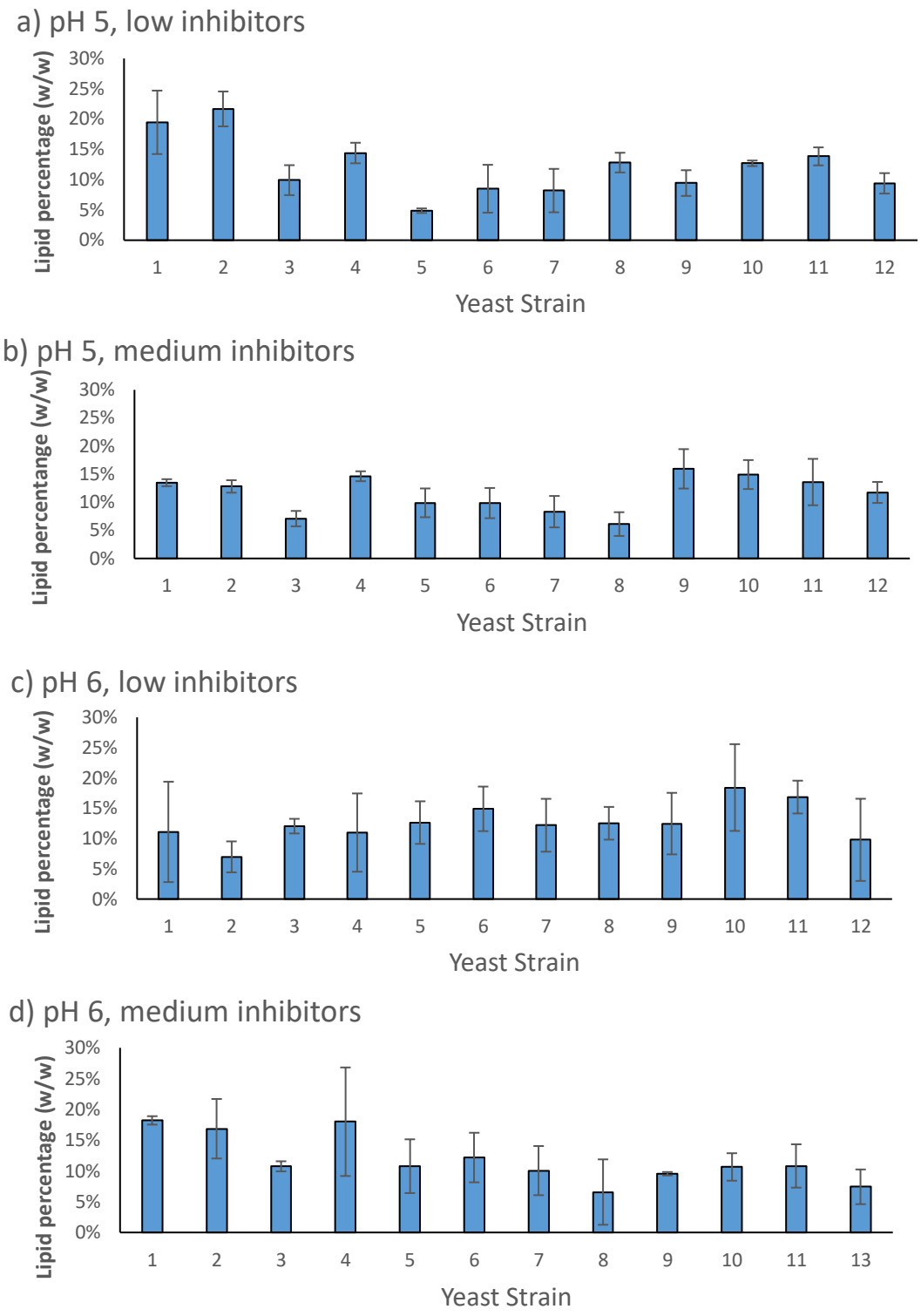


Figure 3.12: a) Lipid extraction for the 12 strains (w/w%) in low inhibitors content, pH 5 and temperature of 20° b) Lipid extraction for all 12 strains (w/w%) in low inhibitors content, pH 5 and temperature of 20° c) Lipid extraction for all 12 strains (w/w%) in medium inhibitors content, pH 6 and temperature of 20° d) Lipid extraction for all 12 strains (w/w%) in medium inhibitors content, pH 6 and temperature of 25°.

None of the strains selected produced high levels of oil under the conditions examined, 20& lipid production is the standard for yeasts to be considered oleaginous. But the four *metschnikowia* strains (1-4) showed the highest lipid production (15-20%). All three are from the *Metschnikowia aff. chrysoperlae* which are closely related to *M. pulcherrima*.

3.3.4 Lipid profile

All lipid samples were transesterified and analyzed using Gas Chromatography-Mass Spectrometry (GC-MS), they were compared against known standards. A representative chromatograph is shown in figure 3.13 with the labelling of the fatty acids.

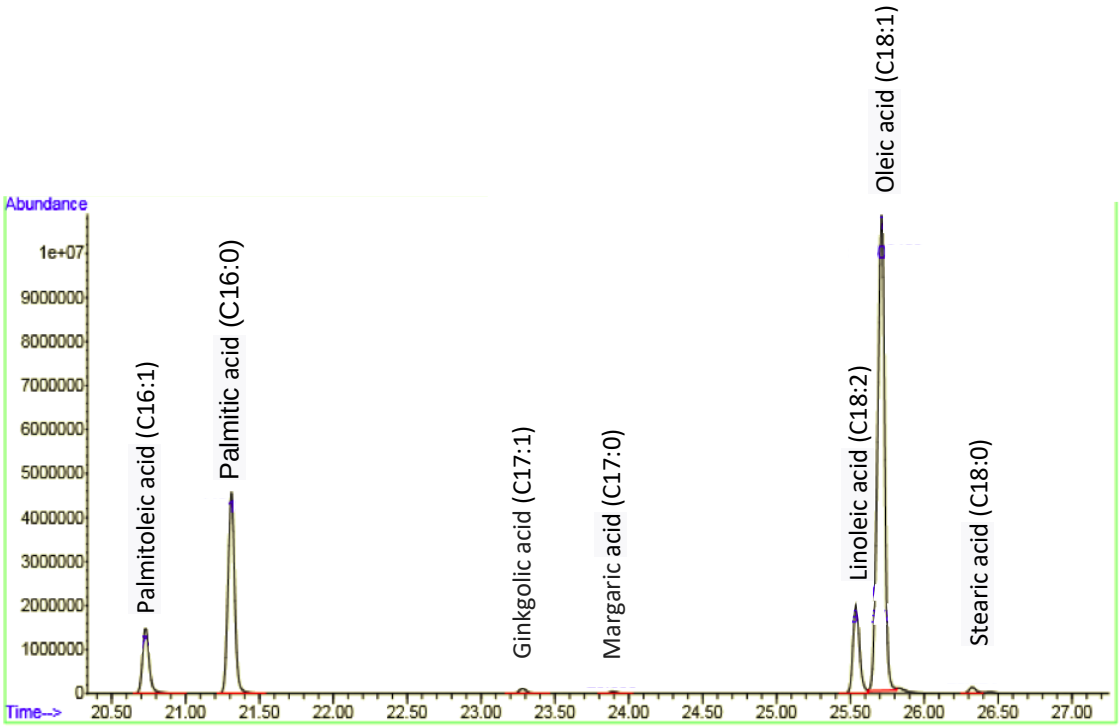


Figure 3.13: An example of GC-MS peaks of the fatty acids produced in the transesterification of oil extracted from yeasts.

Almost all oleaginous yeast contains a simple fatty acid profile, with four main fatty acids, palmitic acid (16:0), stearic acid (18:0), oleic acid (18:1) and linoleic acid (18:2) as summarised in Table 3.4. In the vast majority of literature reports, oleaginous yeasts tend to produce the monounsaturated component, oleic acid in the highest abundance. This gives the majority of yeast oils a similar composition to rapeseed oil. There are a handful of examples however, where the lipid profile is more akin to palm oil [26]. All the lipids produced were analysed, are presented in figures 3.16 – 3.27 below.

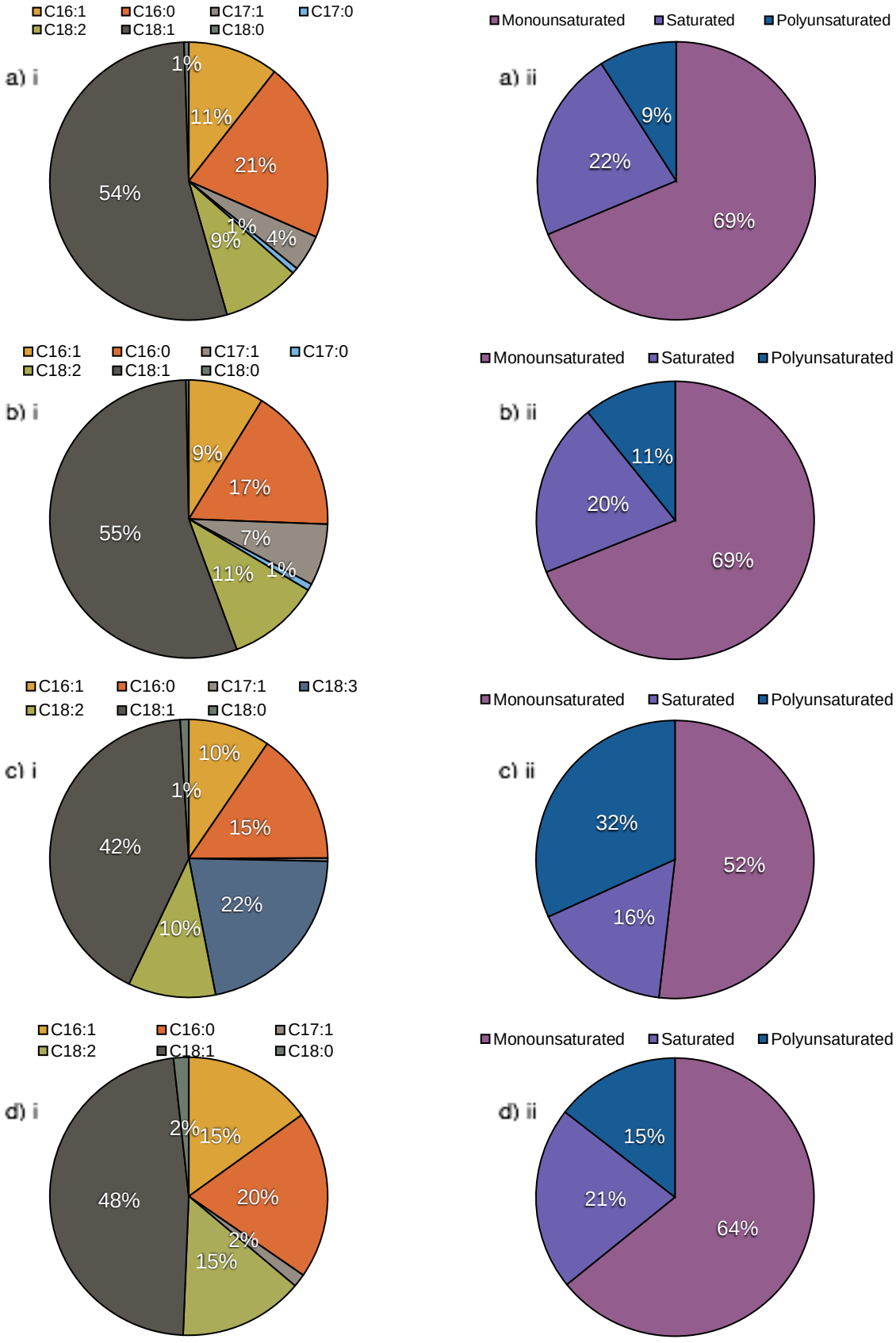


Figure 3.14: Lipid profile for strain no. 1 : *Metschnikowia aff. chrysoperlae* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, at 20°C; b) Low Inhibitors, pH 5, at 25°C; c) Medium Inhibitors, pH 6, at 20°C; d) Medium Inhibitors, pH 6, at 25°C

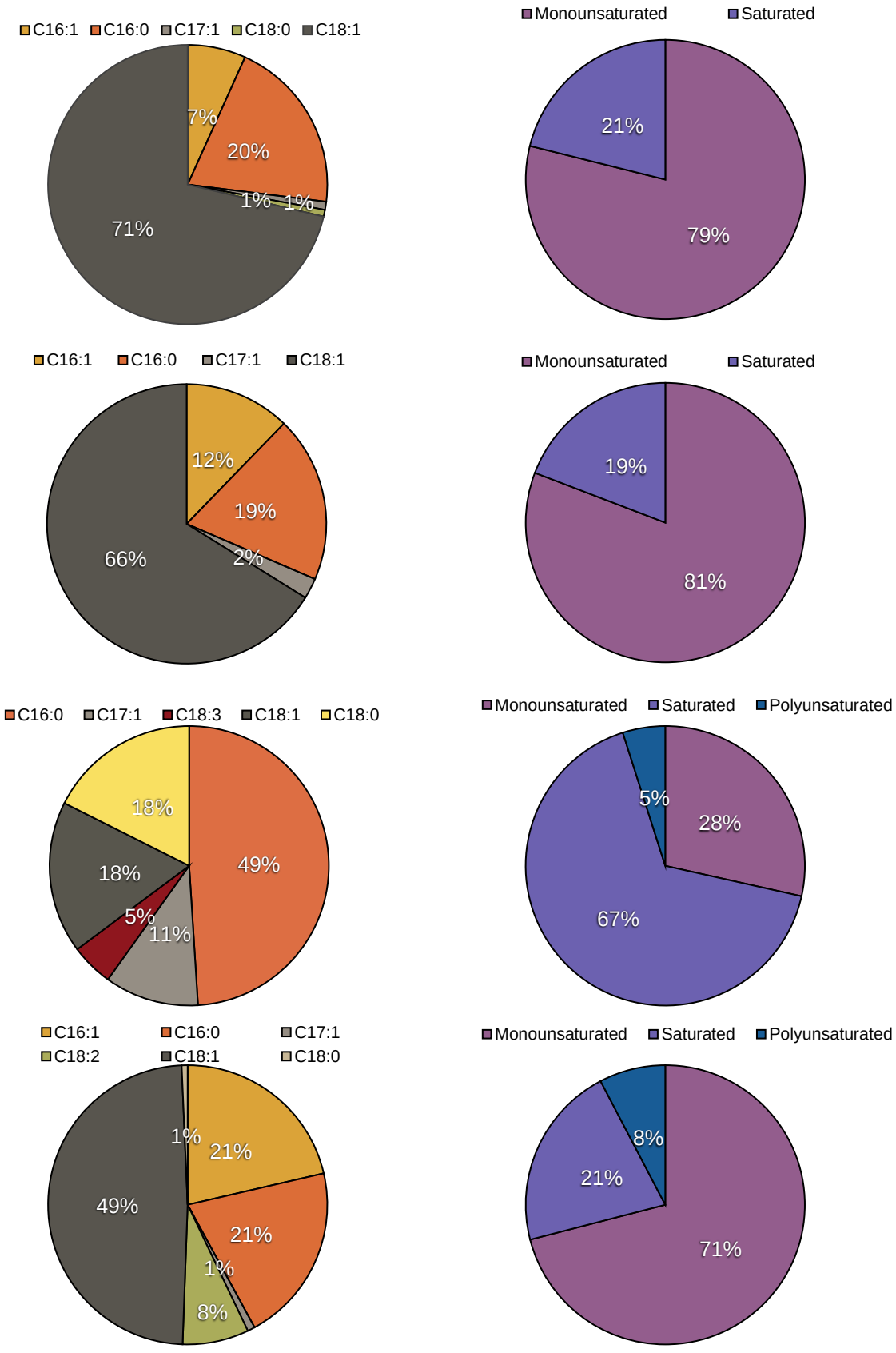


Figure 3.15: Lipid profile for strain no. 2 : *Metschnikowia aff. chrysoperlae* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, at 20°C; b) Low Inhibitors, pH 5, at 25°C; c) Medium Inhibitors, pH 6, at 20°C; d) Medium Inhibitors, pH 6, at 25°C

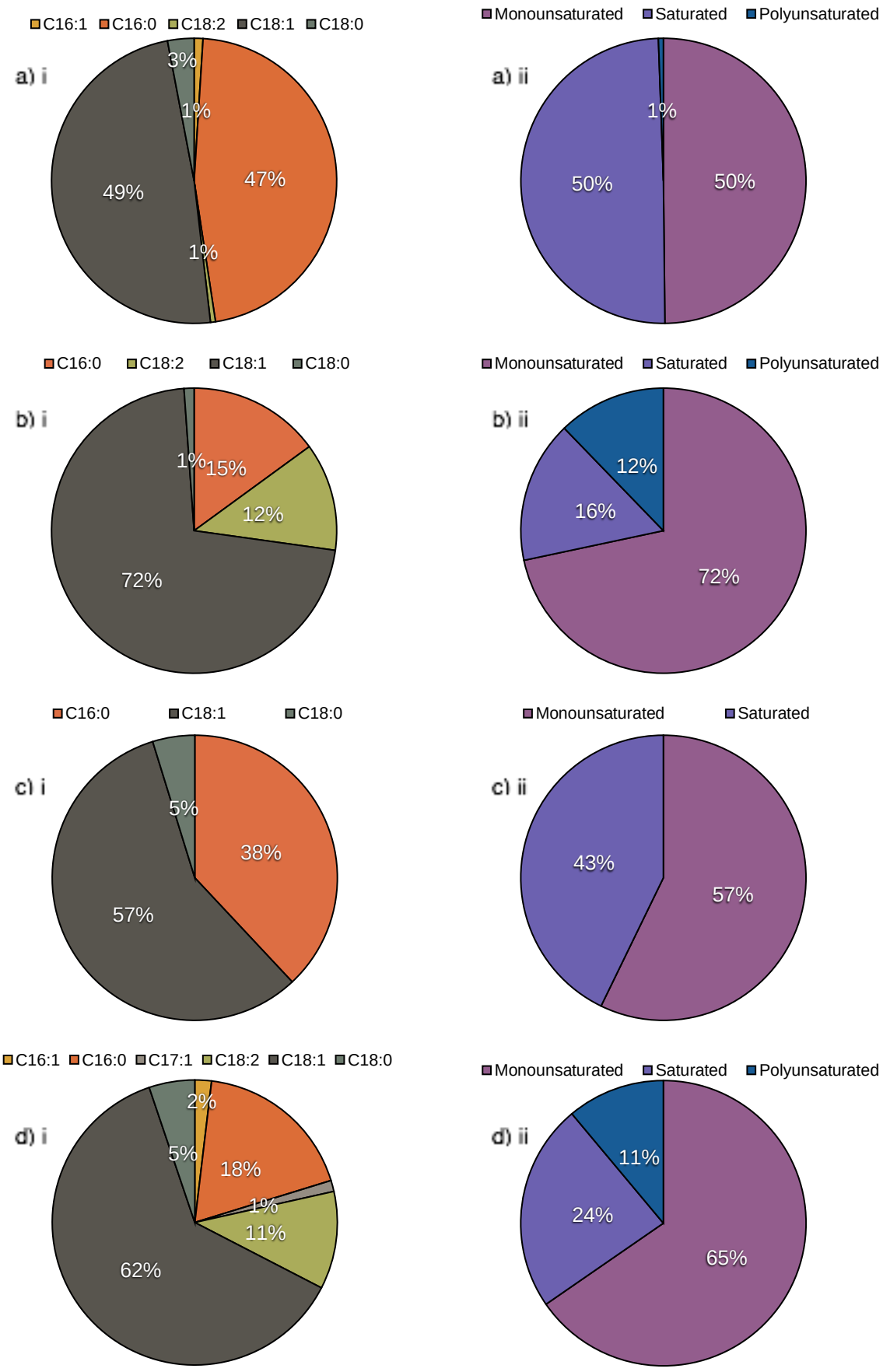


Figure 3.16: Lipid profile for strain no. 3 : *Metschnikowia aff. chrysoperlae* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, at 20°C; b) Low Inhibitors, pH 5, at 25°C; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

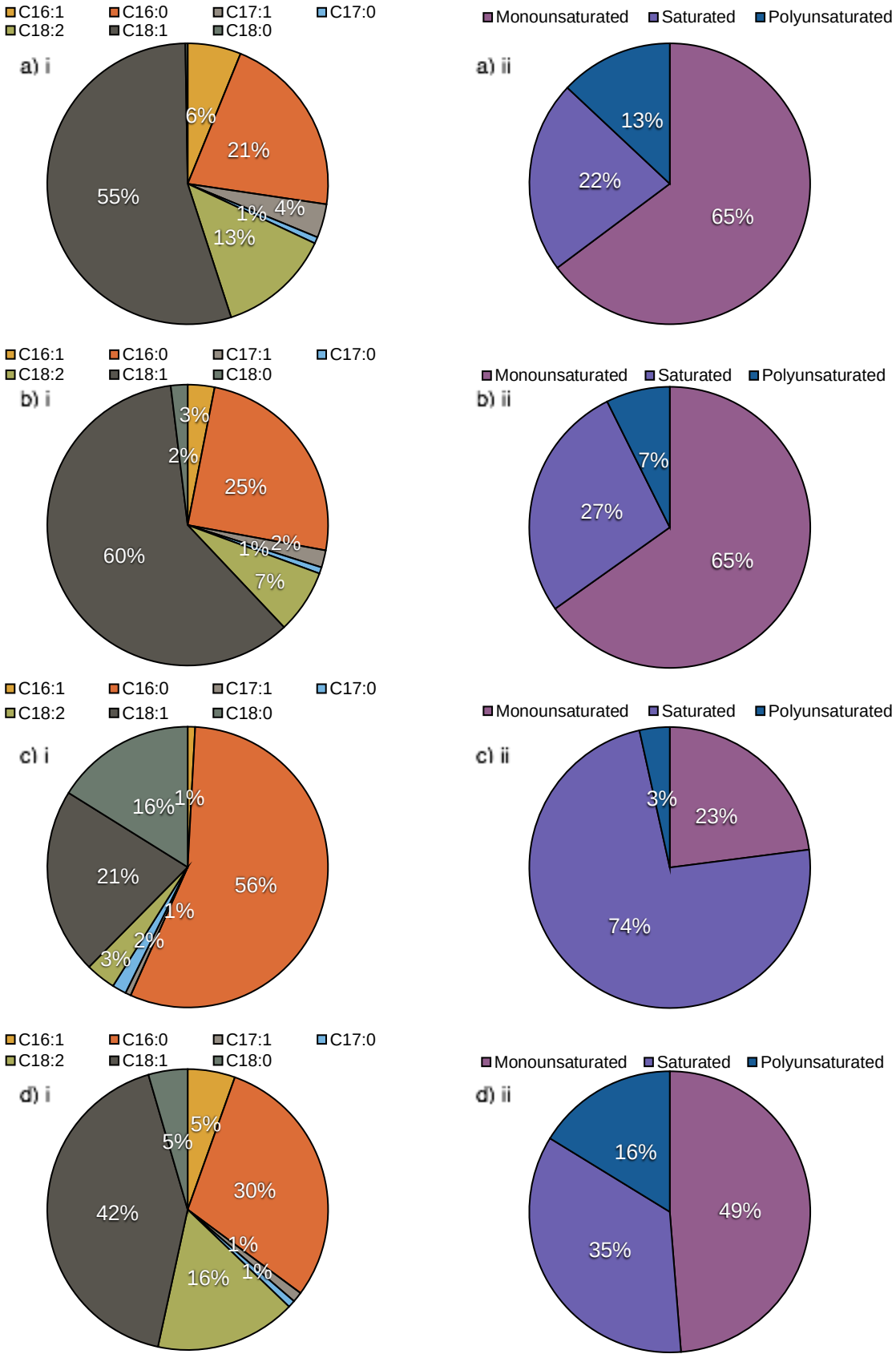


Figure 3.17: Lipid profile for strain no. 4 : *Metschnikowia aff. Chrysoperlae* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, at 20°C; b) Low Inhibitors, pH 5, at 25°C; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

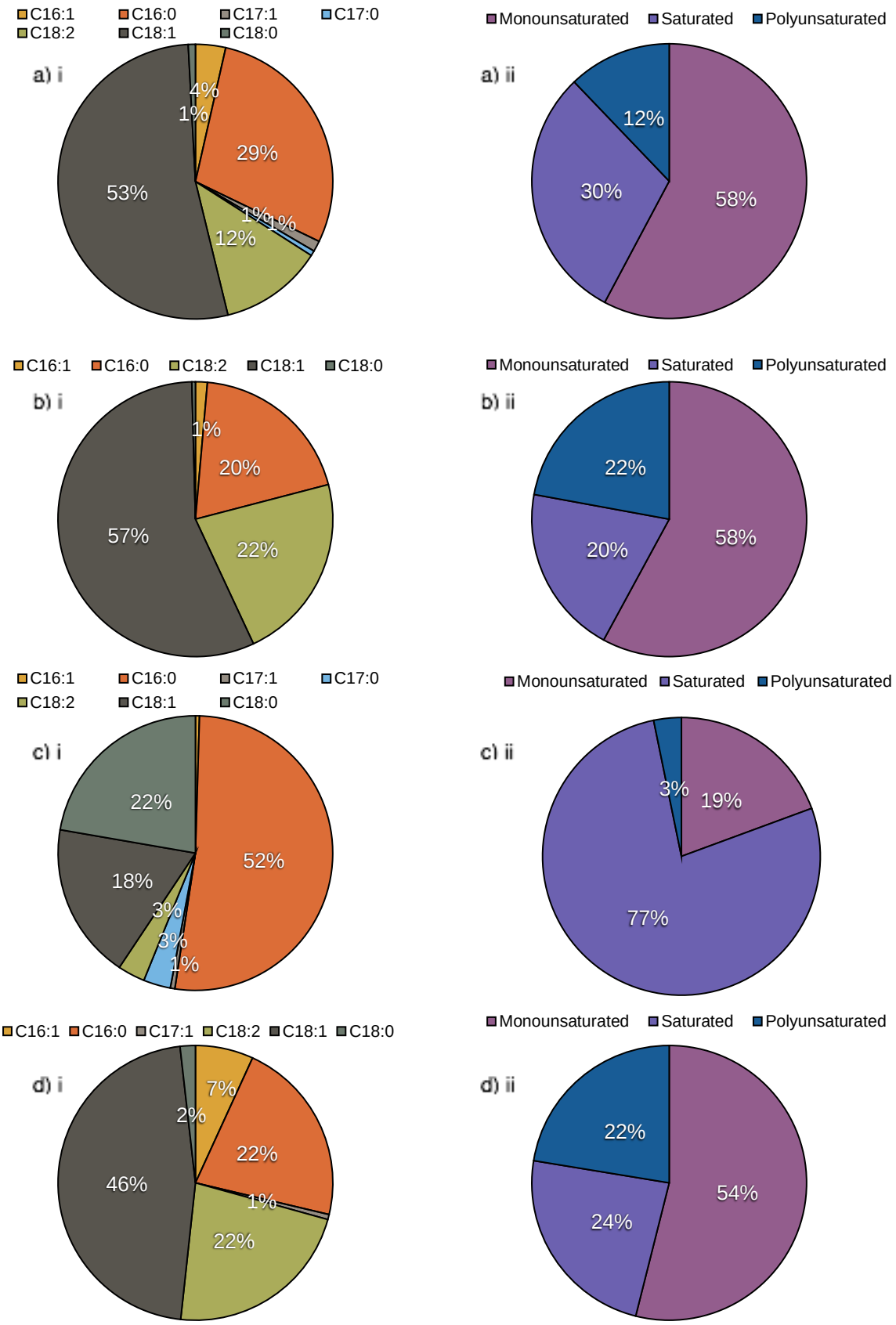


Figure 3.18: Lipid profile for strain no. 5 : *Metschnikowia pulcherrima* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

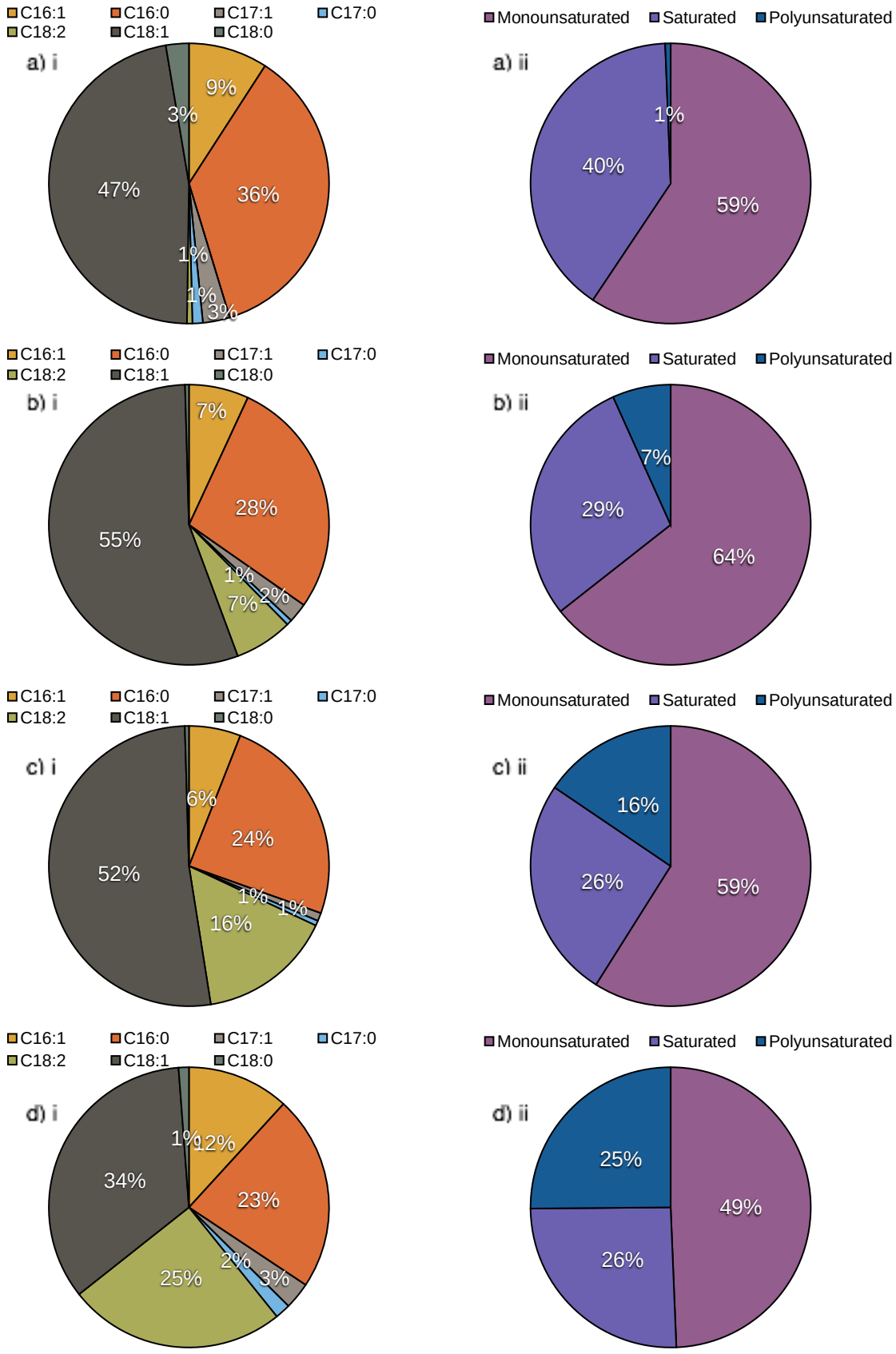


Figure 3.19: Lipid profile for strain no. 6 : *Metschnikowia pulcherrima* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

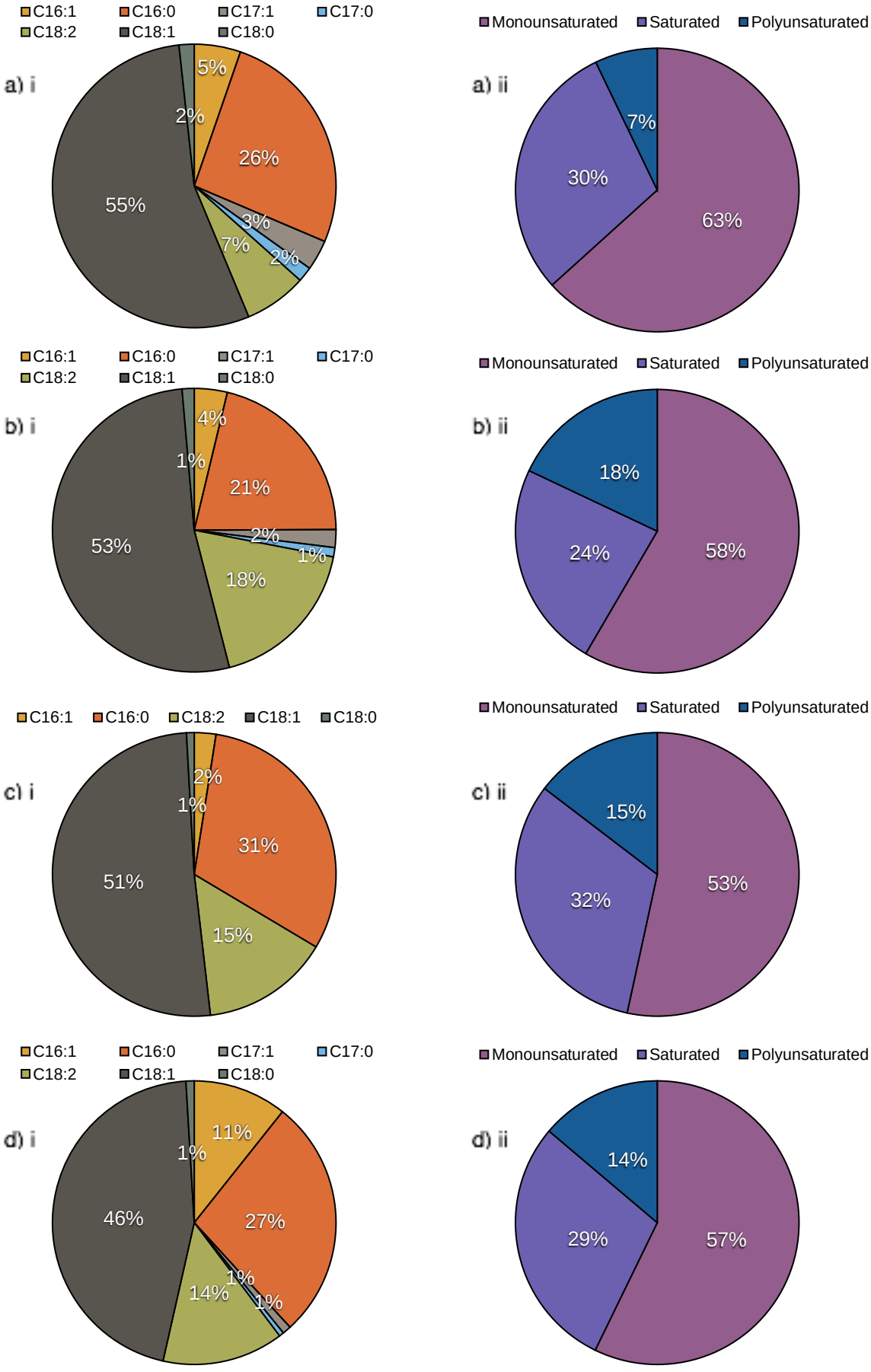


Figure 3.20: Lipid profile for strain no. 7 : *Metschnikowia pulcherrima* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

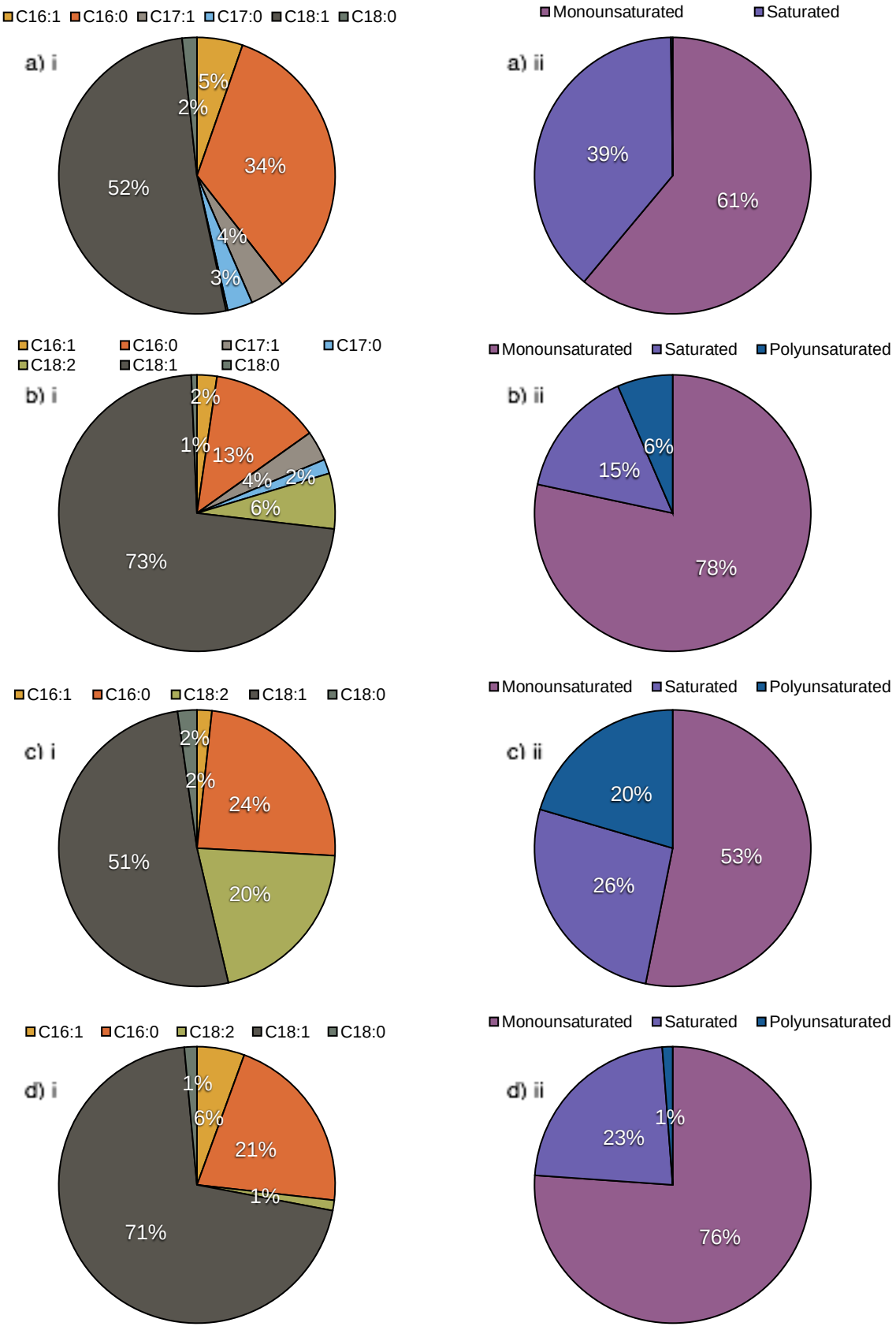


Figure 3.21: Lipid profile for strain no. 8 : *Candida friedrichii* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

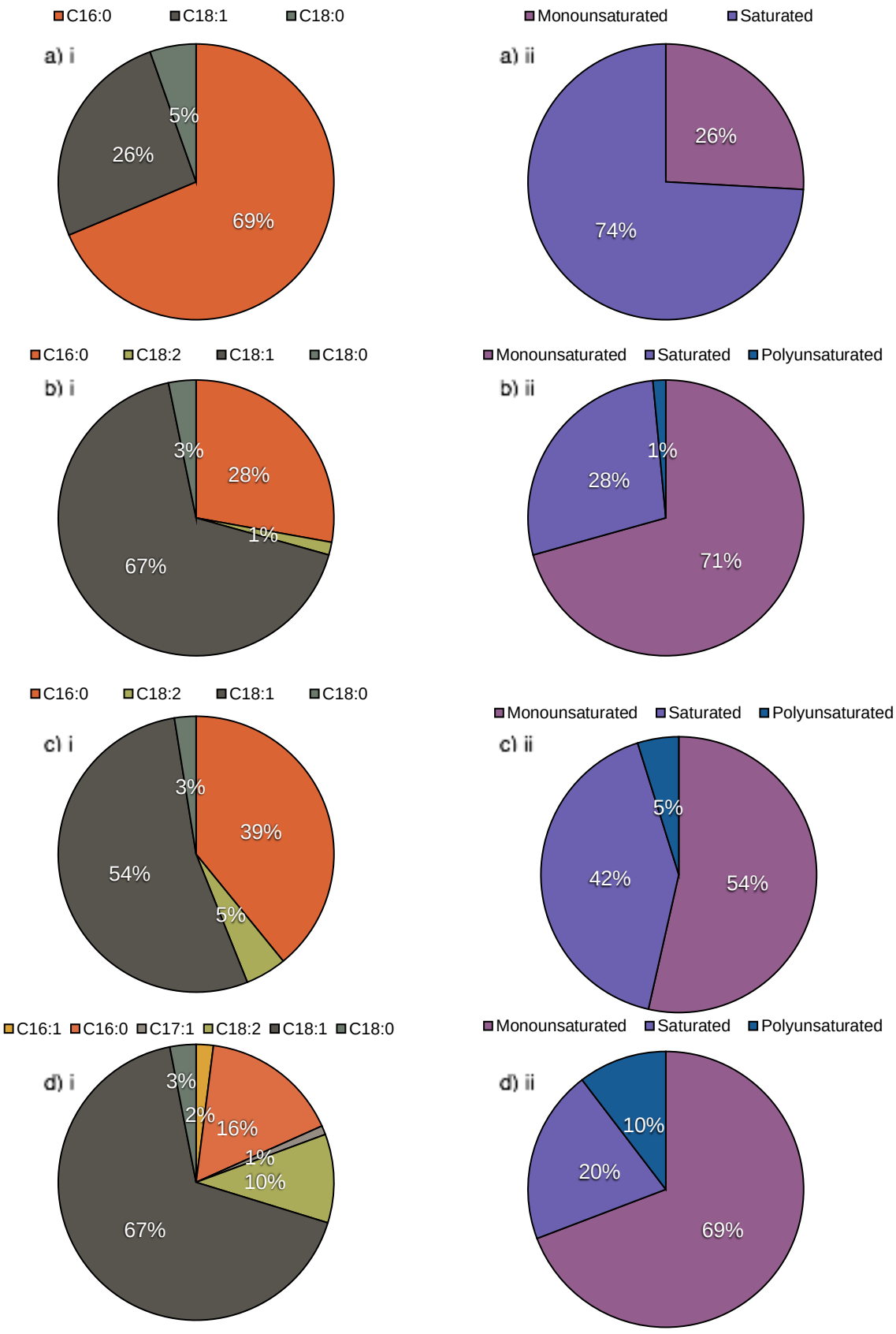


Figure 3.22: Lipid profile for strain no. 9 : *Candida oleophila i* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

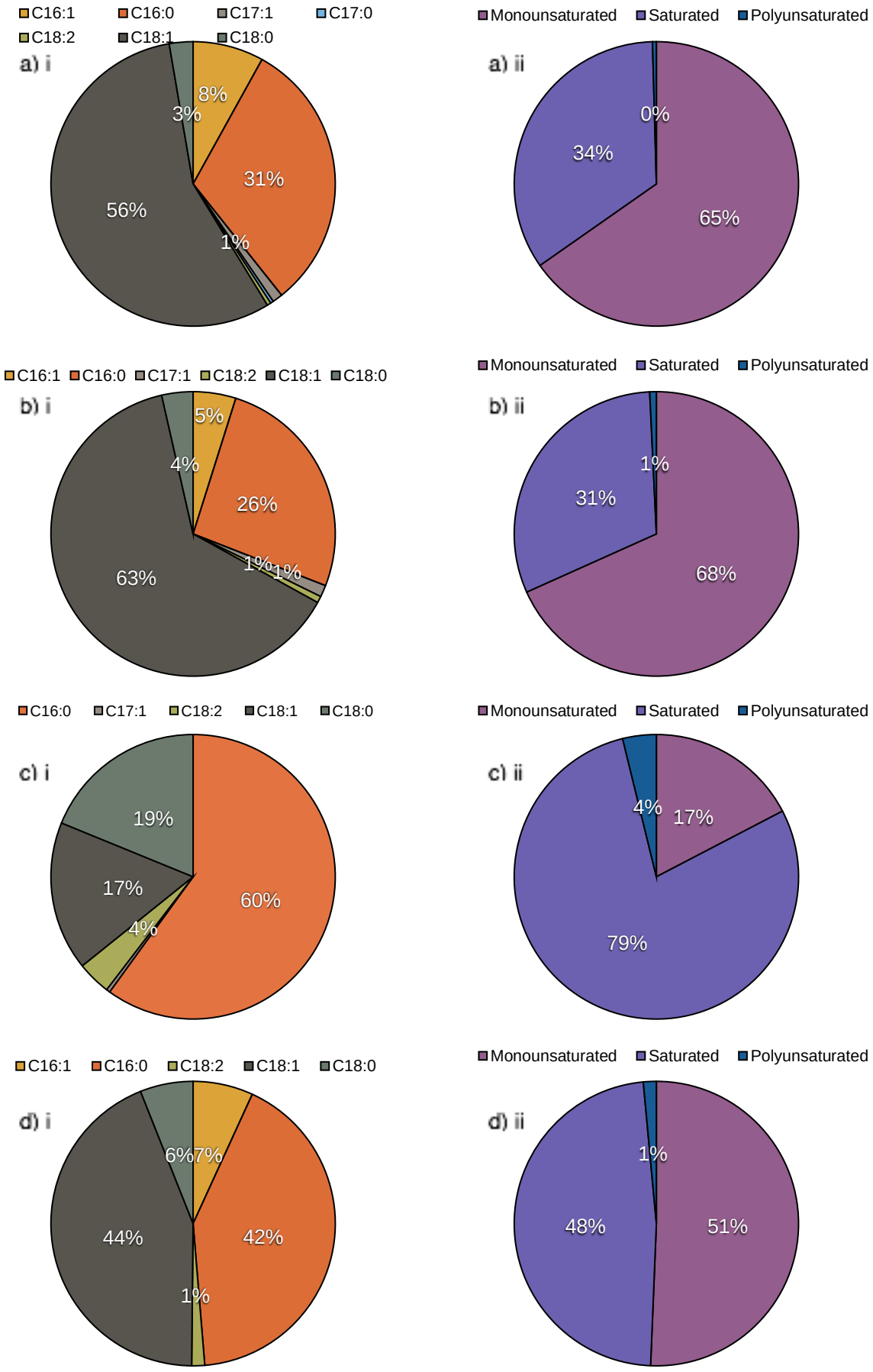


Figure 3.23: Lipid profile for strain no. 10 : *Meyerozyma guilliermondii* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

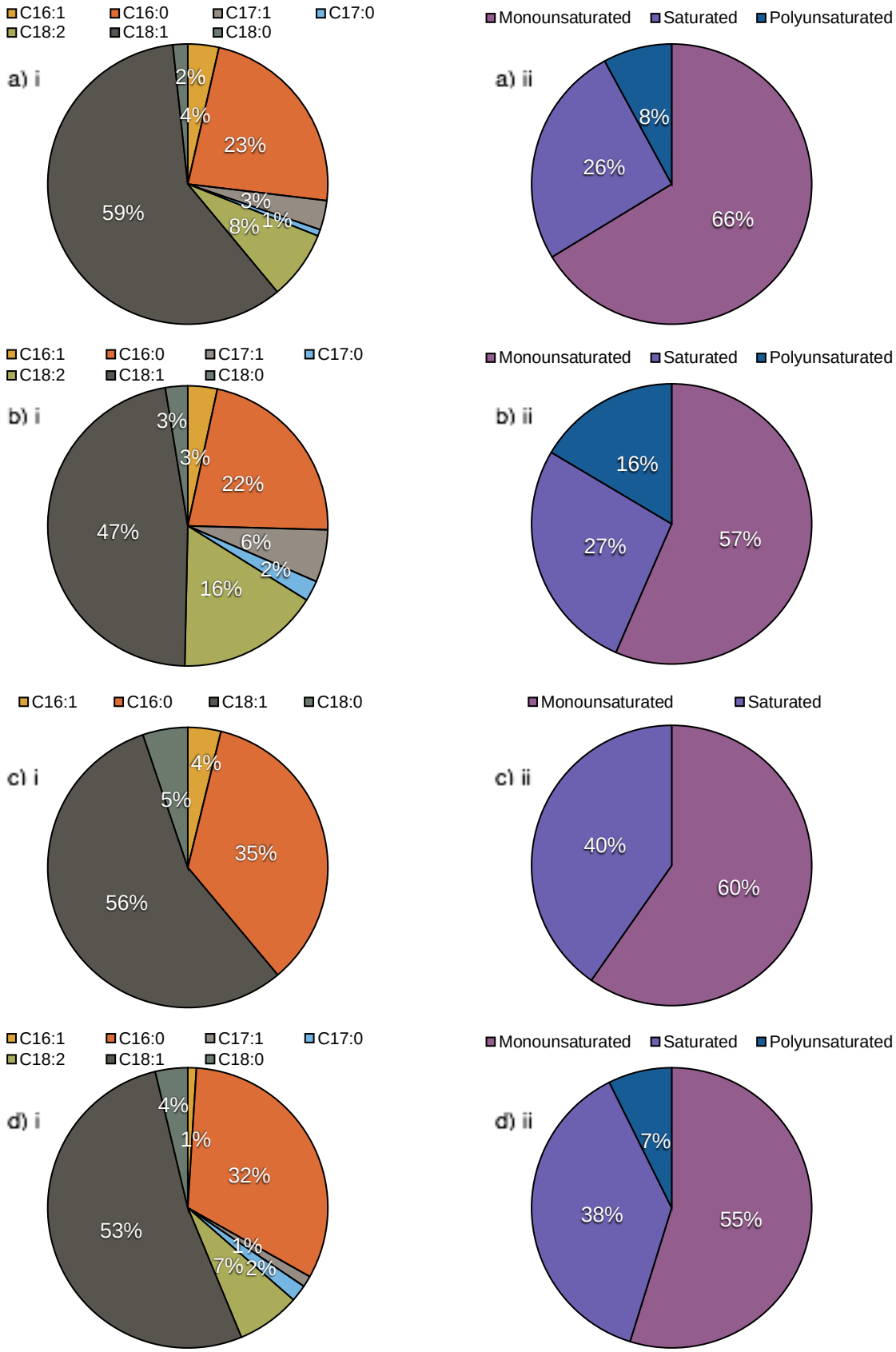


Figure 3.24: Lipid profile for strain no. 11 : *Wickerhamomyces anomalus* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH

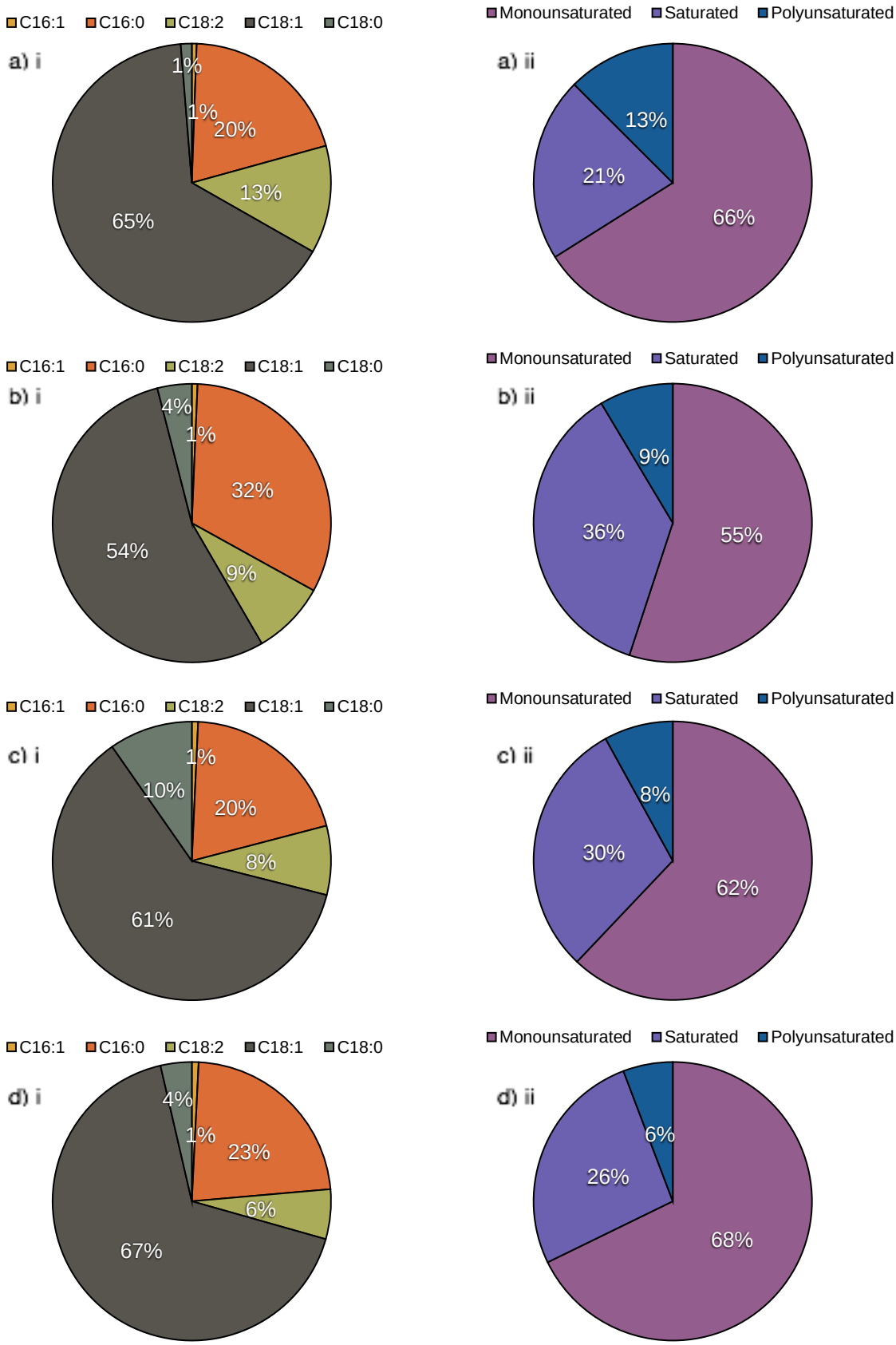


Figure 3.25: Lipid profile for strain no. 13 : *Rhodotorula mucilaginosa* cultured in Yeast Minimal Medium in a) Low Inhibitors, pH 5, and at 20°C; b) Low Inhibitors, pH 5, and at 25°C ; c) Medium Inhibitors, pH 6, and at 20°C; d) Medium Inhibitors, pH 6, and at 25°C.

Strain 1, *Metschnikowia aff. Chrysoperlae*, has a typical profile for an oleaginous yeast, and in all different conditions of growth, produces oleic acid (C18:1) in higher amounts than the other lipids, accumulating higher than 50% Monounsaturated Fatty Acid (MUFA). The temperature did not change the lipids substantially, though increasing the pH did reduce the MUFA content.

This was also observed for the second *M. aff. Chrysoperlae* strain, where at pH 5 the vast majority of the lipid is monounsaturated, while this changed at pH 6, the yeast predominantly produced less saturates and more polyunsaturated esters.

Another *M. aff. Chrysoperlae* strain 3, also demonstrated this type of lipid profile, however, despite the monounsaturated component being around 50%, where the saturates higher, at the low temperature and pH then in pH 6 and 25˚C. The lipid product is a close fit for palm oil. The lipid profile changes dramatically over the other conditions, as well as for the final *M. aff. Chrysoperlae* strain 4. This demonstrates that for this specific yeast, the pH, inhibitor concentration strain type and temperature all have a large effect on the lipid profile. So while to maximise growth the pH and inhibitor concentration would need to be balanced, this would have the effect of changing the lipid profile, meaning a guaranteed constant lipid profile could not be maintained.

Strain 5, 6 and 7 were classified as *Metschnikowia pulcherrima.* Strain 5, at pH 5, produced an oil high in MUFAs, oil from pH 6 and medium inhibitors at 25°C, showed a profile closer to palm oil, while the oil at pH 5, low inhibitors and 25 °C, produced a very high saturated oil. Both strain 6 and 7 showed a similar trend with a product akin to rapeseed oil at pH 5, though at pH 6 showed an inclination towards a palm oil like product.

Strain no. 8 and 9 were two Candida species. The first, *Candida friedrichii*, at lower temperature produced a percentage of monounsaturated ester closer to palm oil. While at 25°C, the lipid profile with more than 75% monounsaturated - closer to of safflower oil. *Candida oleophila,* strain no. 9, demonstrated a larger lipid temperature effect. With the lipids at 20 °C being significantly saturated with the higher percentage are of C16:0, irrespective of the other conditions, in contrast to the high temperature

where the oil closely fits the rapeseed oil profile. This was the most saturated product out of all the yeasts examined.

Strain no. 10, *Meyerozyma guilliermondii*, at pH 5 and low inhibitors, produced an oil closer to rapeseed oil. At pH 6 and medium inhibitors, showed a lipid profile similar to palm oil. *Wickerhamomyces anomalus*, strain no. 11 produced oil with higher levels of monounsaturates. The last strain, *Rhodotorula mucilaginosa*, produced an oil with a similar profile to rapeseed oil, except for the ones cultured in pH5, low inhibitors and 25°C.

Table 3.4 below summarizes the lipid profiles of all the 12 different stains in all four different conditions.

Strain No. (conditions)		Saturated Fatty Acid (%)			Monounsaturated Fatty Acid (%)			Polyunsatirated Fatty Acid (%)	
		C 16:0	C17:0	C18:0	C 16:1	C17:1	C18:1	C18:2	C18:3
1	(a)	20.983	0.693	0.553	10.580	4.244	53.886	9.060	—
	(b)	16.797	0.791	0.329	8.793	7.173	55.290	10.826	—
	(c)	15.412	—	0.994	9.549	0.372	41.925	10.187	21.561
	(d)	19.530	—	1.764	15.067	1.535	47.584	14.520	—
2	(a)	20.184	0.165	0.780	6.731	0.941	71.199	—	—
	(b)	19.202	—	—	12.258	2.389	66.152	—	—
	(c)	48.993	—	17.596	—	10.897	17.596	—	4.918
	(d)	20.659	—	0.686	21.383	0.868	48.734	7.670	—
3	(a)	46.600	—	2.996	0.998	—	48.839	0.566	—
	(b)	14.967	—	1.134	—	—	71.662	12.237	—
	(c)	38.026	—	4.773	—	—	57.201	—	—
	(d)	18.351	—	5.184	1.886	1.263	62.231	11.084	—
4	(a)	21.186	0.750	0.262	6.148	3.890	54.752	13.012	—
	(b)	24.803	0.735	1.931	3.085	1.983	60.110	7.354	—
	(c)	55.807	1.628	16.127	0.853	0.644	21.469	3.472	—
	(d)	29.677	0.839	4.515	5.459	1.182	42.086	16.243	—
5	(a)	28.618	0.607	0.839	3.546	1.266	52.969	12.155	—
	(b)	19.546	—	0.419	1.397	—	56.522	22.116	—
	(c)	51.926	3.203	22.274	0.472	0.532	18.380	3.211	—
	(d)	21.870	—	1.792	6.831	0.601	46.492	22.414	—
6	(a)	36.149	1.201	2.650	9.144	3.112	47.120	0.624	—
	(b)	27.794	0.628	0.445	6.908	2.272	55.222	6.730	—
	(c)	24.479	0.585	0.471	5.976	0.939	52.030	15.520	—
	(d)	22.544	1.783	1.183	11.786	3.098	34.468	25.138	—
7	(a)	26.112	1.726	1.722	5.287	3.465	54.555	7.133	—
	(b)	21.170	1.063	1.376	3.759	2.039	52.617	17.975	—
	(c)	31.082	—	0.876	2.466	—	50.948	14.628	—
	(d)	27.493	0.540	0.935	10.697	0.990	45.525	13.820	—
8	(a)	34.047	2.939	1.726	5.359	4.070	51.651	0.208	—
	(b)	12.835	1.683	0.659	2.326	3.511	72.501	6.485	—
	(c)	24.131	—	2.250	1.737	—	51.407	20.475	—
	(d)	21.221	—	1.464	5.558	—	70.548	1.209	—
9	(a)	68.697	—	5.419	—	—	25.884	—	—
	(b)	27.837	—	3.221	—	—	67.445	1.498	—
	(c)	39.112	—	2.526	—	—	53.536	4.826	—
	(d)	16.291	—	3.061	2.017	1.050	67.189	10.393	—
10	(a)	31.244	0.340	2.696	8.057	1.169	56.055	0.439	—
	(b)	26.061		3.555	4.861	1.339	63.434	0.750	—
	(c)	59.987		18.828		0.422	16.942	3.821	—
	(d)	41.854		6.027	6.836		43.812	1.471	—
11	(a)	23.305	0.752	1.709	3.583	3.391	59.320	7.940	—
	(b)	22.054	2.378	2.576	3.393	6.053	47.077	16.470	—
	(c)	35.090		5.191	3.815		55.904		—
	(d)	32.179	1.938	3.734	0.988	1.295	52.521	7.345	—
13	(a)	20.158		1.273	0.556		65.494	12.519	—
	(b)	32.380		3.987	0.660		54.393	8.580	—
	(c)	20.205		9.699	0.726		61.355	8.014	—
	(d)	22.829		3.618	0.787		66.996	5.770	—

Table 3.4: A summary of the lipid profile for each fermentation experiment cultured in the four different conditions: (a) Low Inbibitors, ph 5, 20°C; (b) Low Inbibitors, ph 5 , 25°C (c) Medium Inhibitors, pH 6, 20°C (d) Medium Inhibitors, pH 6, 25°C. The fatty acids are attained are namely Palmitic acid (C16:0), Palmitoleic acid (C16:1), Margaric acid (C17:0), Ginkgolic acid (C17:1), Stearic acid (C18:0), Oleic acid (C18:1), Linoleic acid (C18:2), γ-Linolenic acid (18:3).

3.4 Conclusions

The strains of yeast isolated from the previous chapter were cultured with alternative model hydrolysates, over a range of pHs and inhibitor loadings. Under these conditions all the yeasts tended to grow similarly with no large differences between species. With no inhibitors the yeasts could grow at a low pH effectively, however, as the inhibitor loading was increased the optimal pH increased. The differences of glucose and xylose contents of each different lignocellulose waste simulation does not give much effect to the growth performance of all 12 strains. This prima facie provides an advantage of flexibility in the ability to utilise any form of lignocellulose waste as a carbon source in the industrial settings.

All of the strains in this collection could produce the long chain fatty acids that made up vegetable oils, though under the inhibitor rich conditions only the yeast in the *Metschnikowia* genus could produce over 20% of lipid content in 5 days. No one strain showed a superior production of lipid.

The majority of the yeasts produced a highly monounsaturated oil under the range of conditions examined, though a number of the yeasts showed the ability to produce a palm oil type product under certain conditions. This demonstrated that there is a strong inherent trait for vegetable oil-like lipid by these resilient strains of yeasts. However, the inconstancy of the lipid profile demonstrates that with a variable feedstock and processing conditions, no specific type of lipid can be guaranteed without either further genetic manipulation or other type of culture control. There is a certain level of genetic control to the production of lipids aside that of the inhibitor concentrations, pH conditions & its feedstock. If the understanding of the metabolic, and variability of the medium specifications be done in cohesion, the optimum production of lipid with the consistency that of palm oil could be achieved.

Chapter 4

Draft Genome Sequence of *Metschnikowia pulcherrima*, strain WW02B (NCBI Accession No. FR07618473), an oleaginous yeast with the potential as palm oil alternative producer.

4.1 Introduction

In the earlier chapter, we have established the identification of the 12 studied strains by using Internal Transcribed Spacer 1 and 2 Regions for Rapid Detection and Identification of yeast species [128, 129]. Here, only a small identification marker were used to be amplified & the sequenced matched with what is on the database for yeast identification [130]. The range of database for identification of yeasts and other organism are shared on the virtual platform in an unlimited extent. Identification is but the surface of the genetics information being widely shared in the bioinformatics world. The increasing accumulation of genetics data and also becoming more important now is the genomic data especially human genome [131].

Genomic sequencing was initiated by scientists on the bacterium *Haemophilus influenzae* Rd in 1995 [132], using the very early shotgun sequencing method [133]. In 1996, the first Eukaryotic organism was fully sequenced, namely the "Brewer's yeast", *Saccharomyces cerevisiae*, through an international collaboration, where an extensive genetic map that defines the haploid set of 16 chromosomes [134]. The more efficient techniques constructed from the de novo sequencing of this yeast had led the burst of many more genomic sequencing of more complex Eukaryotes including human in 2001 [135]. Other species genome sequenced after the onset of the first Eukaryote are: a nematode (Caenorhabditis elegans) (1998) [136]; an insect (Drosophila melanogaster) (2000) [137]; a dicotyledonous plant (Arabidopsis thaliana) (2000) [138].

The completion of genome sequencing of the model fungus *S. cerevisae*, sequencing of large numbers of fungal genomes followed suit. Sequencing of large numbers of fungal genomes allowed us to recognise the variety of genes encoding enzymes, and pathways that produces several then newly found compounds [139]. Even though the yeasts are very diverse in characteristics, their basic cellular physiology and genetics hold some universal elements with plants and animal cells. These include multicellularity, cytoskeletal structures, cell cycle, circadian rhythm, intercellular signalling, sexual reproduction, development and differentiation [140, 141].

In the *S. cerevisiae* genome project showed the potential and limitations of using homology studies (The basis of searching related genes that have similar sequences and so a new gene can be discovered because of its similarity to an equivalent, already sequenced, gene from a different organism). The yeast genome contains approximately 6000 genes, 30% had been identified by established analysis, while the remaining 70% were studied by homology analysis [134, 142]. The result is displayed in the figure 4.1.

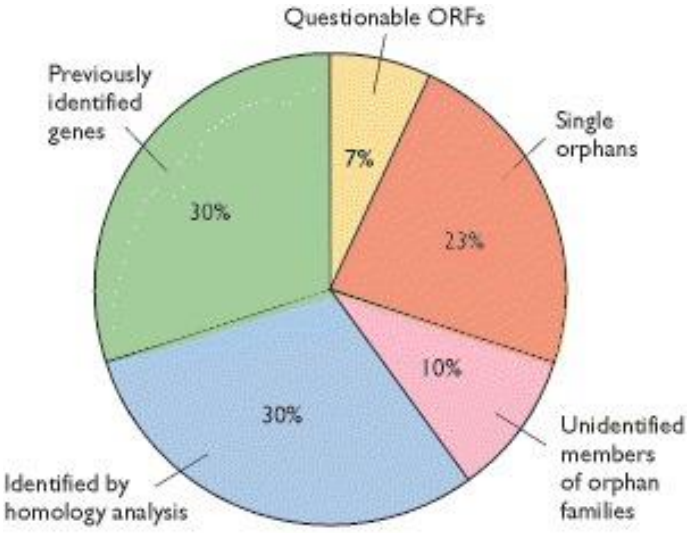


Figure 4.1: Categories of gene in the yeast genome [142]

Because of the simplicity of growing and manipulating it, *S. cerevisiae* is vastly used in the exploration of biochemistry, molecular biology, cell biology and systems biology. Furthermore, it has extensive conservation of its genes and pathways with those of higher organisms. Many modern genome-wide biotechnologies, for example, the creation of bar-coded systematic deletion sets; large-scale detection of protein–protein and genetic interactions and subcellular localizations; transcriptome, proteome and metabolome analysis, were original developed using yeast before being more widely applied to other organisms [143].

With the latest advancements in biotechnology, from lab procedures to complex machineries, an overflow of genomic sequences of many genera have been produced. And this provide a strong platform for comparative genome analysis. Yeasts and fungi are ideal organisms for comparative genomic studies in eukaryotes due to their small and compact genomes and because they include a number of species, that have been, and continued to be, studied extensively in genetics [144]. Comparative genome analysis of related species presents a formidable and general approach for identifying functional elements without previous knowledge of the function. Evolution persistently interferes with genome sequence and proves the results by natural selection, where it provides elements that by virtue clearly having a superior scale of conservation across connected species. The advantage of this is that its strength can be improved by increasing the number of species being researched [145].

Metschnikowia pulcherrima and *S. cerevisiae* are Ascomycete yeasts (phylum Ascomycota: subphylum Saccharomycotina: class Saccharomycetes: order Saccharomycetales [146]. *M. pulcherrima* is a post-harvest control agent recognised as to severely limit the growth of other microbes in the natural environment [147], The yeast succeeds in the must even though it is high in acidity (pH 3 to 4) and sugar concentration (>100 g/L), and the cultures of *M. pulcherrima* do not require iron or vitamins other than biotin [148, 149]. We have also established in the earlier chapters that *M. Pulcherrima* can withstand medium strength of mixture of inhibitors produced from lignocellulosic degradation. Other than that it can also utilizes many different type of saccharides efficiently. Hence, a very resilient strain for usage in industries.

We have chosen the strain tagged as WW02B collected from the bioprospecting collection. It was attained from some Blackberries (*Rubus fruticosus* agg.) in the bushes between the Westwood students' accommodation and the peripheral of Bath Golf Club at the coordinate: 51°22'52.8"N 2°19'57.6"W. Using the ITS primers, this strain identified as 98% identity to *Metschnikowia pulcherrima* strain NRRL Y-7111 small subunit ribosomal RNA gene, partial sequence; internal transcribed spacer 1, 5.8S ribosomal RNA gene, and internal transcribed spacer 2, complete sequence; and large subunit ribosomal RNA gene, partial sequence (BLAST). The accession no. for this strain is FR07618473 (BLAST, NCBI).

4.2 Results and Discussions

4.2.1 Culturing and DNA extraction

The strain was cultured in 1 Litre Soy broth for 3 days shaken in 25°C. And then the biomass was collected in 50-mL sterile conical plastic tubes. Genomics DNA was isolated using a combination of a few DNA extraction and DNA purification techniques as mentioned in 6.1, which is using the Phenol:Chloroform:Isoamyl Alcohol (25:24:1) as the main chemical of this whole protocol. Six 1.5-mL samples were acquired from the extraction are run on a 0.8% agarose gel for quality check. Large DNA samples were seen from the gel image for the first 5 DNA lanes as corresponding to the ladder. However, the samples 1 & 2, displayed blurred spots at the very end of the lane revealing short nucleotides believed to be RNA contaminations. And sample 6 proved to have very little genomics DNA as it displayed very faint band. Therefore, these samples were excluded from being sent for the next quality accession procedure.

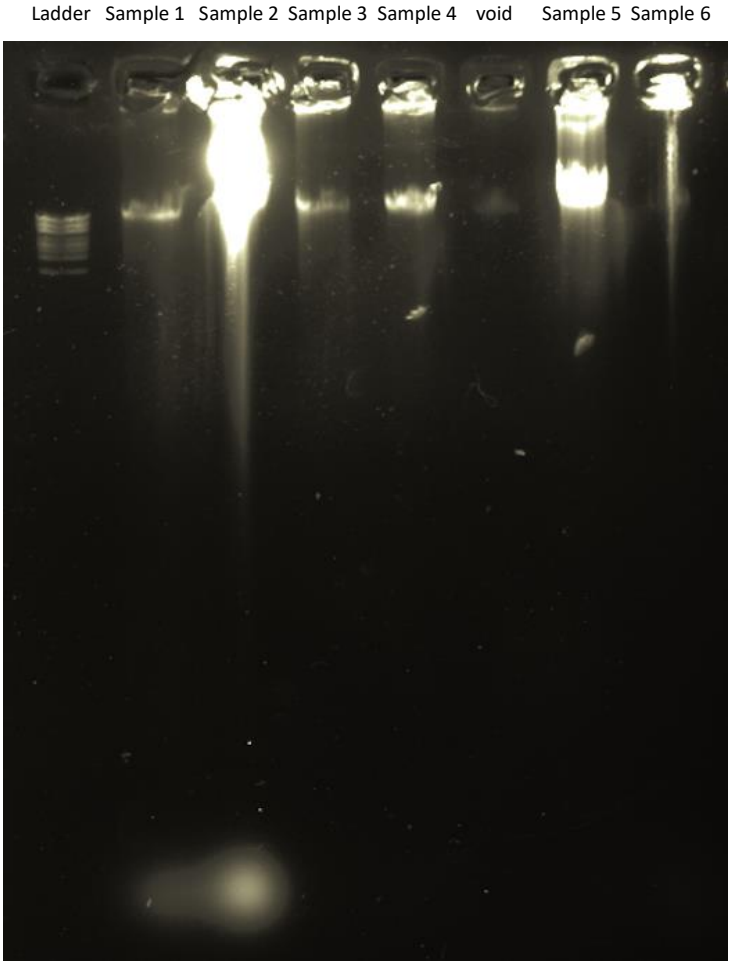


Figure 4.2: Gel electrophoresis of the 6 DNA extraction samples.

The DNA quality of samples 3, 4 & 5 were then accessed in the lab using Qubit® 2.0 fluorometer reading (with QUANT-iT dsDNA assay) and NanoDrop® 1000 UV-Vis Spectrophotometer. The results were as in table 4.1. Only samples 3 met with the requirements of the Exeter Sequencing Service where the DNA sample was send for the Genomic Sequencing procedure.

Table 4.1: The DNA quality as accessed using NanoDrop® 1000 UV-Vis Spectrophotometer and Qubit® 2.0 fluorometer.

	Sample 3	Sample 4	Sample 5	Requirements
NanoDrop reading:				
A260:	0.271 A	0.454 A	1.221 A	
A260/A280:	2.00 A	1.151 A	1.818 A	Between 1.8 and 2.1
A260/A230:	2.078 A	1.020 A	1.693 A	Between 2.0 and 2.2
Concentration:	13.30 ng/ul	18.70 ng/ul	undetecteable	
Qubit reading:	140 ug/ml (1.5 mL)	(high out of range)	180 ug/ml (1.5 mL)	20 ng

4.2.2 Library Preparation and Sequencing

The DNA was fragmented in Covaris E220 before further preparing the library. The DNA sample was prepared for sequencing using NEXTflex® Rapid DNA-Seq Library Prep Kit to construct pair-end libraries. Then, the library prep was accessed for their concentration in Qubit® 2.0 fluorometer. At the end of this part of procedure, a final concentration of 8.80 nM and 3.34 ng/uL was detected, where it was found as acceptable to proceed to genomic sequencing.

The 27.9-Mbp genome of *Metschnikowia pulcherrima* was sequenced using Illumina HiSeq® 2500 Sequencing System. This is the most widely used platform in this biotechnology niche [150]. The Illumina HiSeq 2500 (version 4) provide a versatile

platform to support a wide variety of applications from bacterial and fungal to whole human genome sequencing. The HiSeq has 2 flowcells each containing 8 separate 'lanes' for the high output mode and 2 separate 'lanes' for the rapid mode. The rapid mode can read more bases (250bp) than the high output mode (125bp). They can run completely independently for smaller experiments or in combination for larger genomes (Exeter Sequence Service). In this project, rapid mode, for the sake of time management.

Currently, over 1000 fungal genomes has been sequenced as in the latest (December 2017) list in the NCBI genome page (https://www.ncbi.nlm.nih.gov/genome/). However, the yeasts genome sequencing progress has not been rapid. Thus far, only *M. bicuspidata* and *M. fruticola* from the Metschnikowia genus has been fully sequenced, where and the latest genome, *M. australis* was drafted in early 2017 [151].

We use KRONA in an analysis to illustrate the phylogeny relations between other species from the genetics data in BLAST. Krona is an interactive charting application with objectives are to decode matters concerning the relative abundance of taxa across numerous levels of the hierarchy at the same time. The software complements existing metagenomics visualizations by creating clearer depictions of abundance estimates and by enabling in-depth understanding of the underlying classifications (https://omictools.com/krona-tool).

Figure 4.3 depicts the hierarchy of the higher genes similarity of the strains with the sequenced strain WW02B with other Ascomycota phylum. Having *M. bicuspidata var. bicuspidata* which genome has been sequenced having highest percentages of equivalent genes (27%). In figure 4.4, we went down the hierarchy to the Genus level of Metschnikowia and KRONA reveal that the other Metschnikowia species that have similar genes include *M. rekaufii, M. ziziphicola, M. aff. Chrysoperlae*, and *M. fruticola*, where we have also collected *M. aff. Chrysoperlae* from our bioprospecting sessions. These species are known to be very closely related through a number of earlier studies. None of them has been sequenced genomically. It is deem valuable if it be done for this family, for a thorough genomic comparative studies be perform between these species.

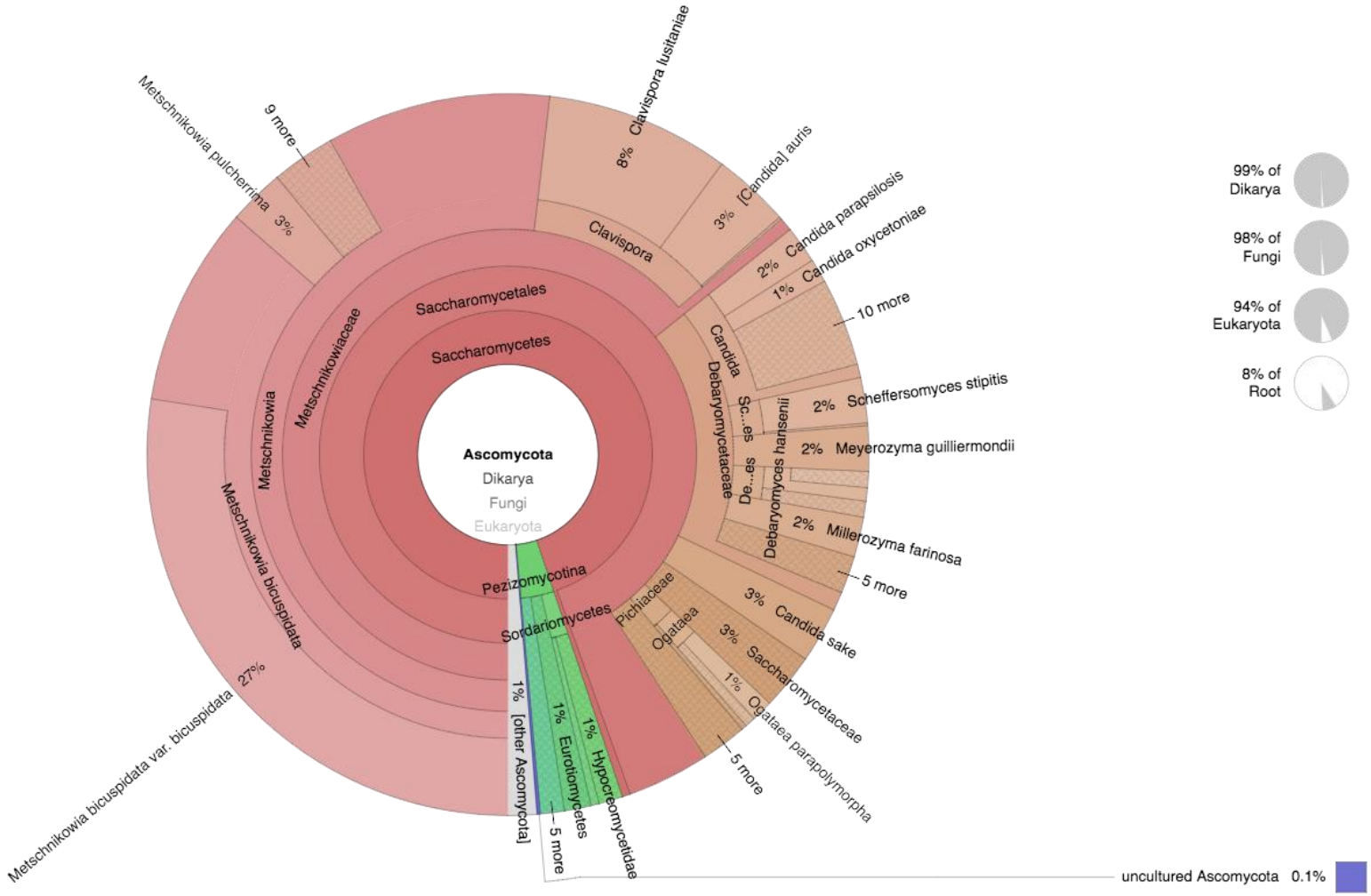


Figure 4.3: KRONA multi-layered pie chart showing the phylogeny relationship of the *M. pulcherrima* strain sequenced with other Ascomycota genome/genetics data from online databases.

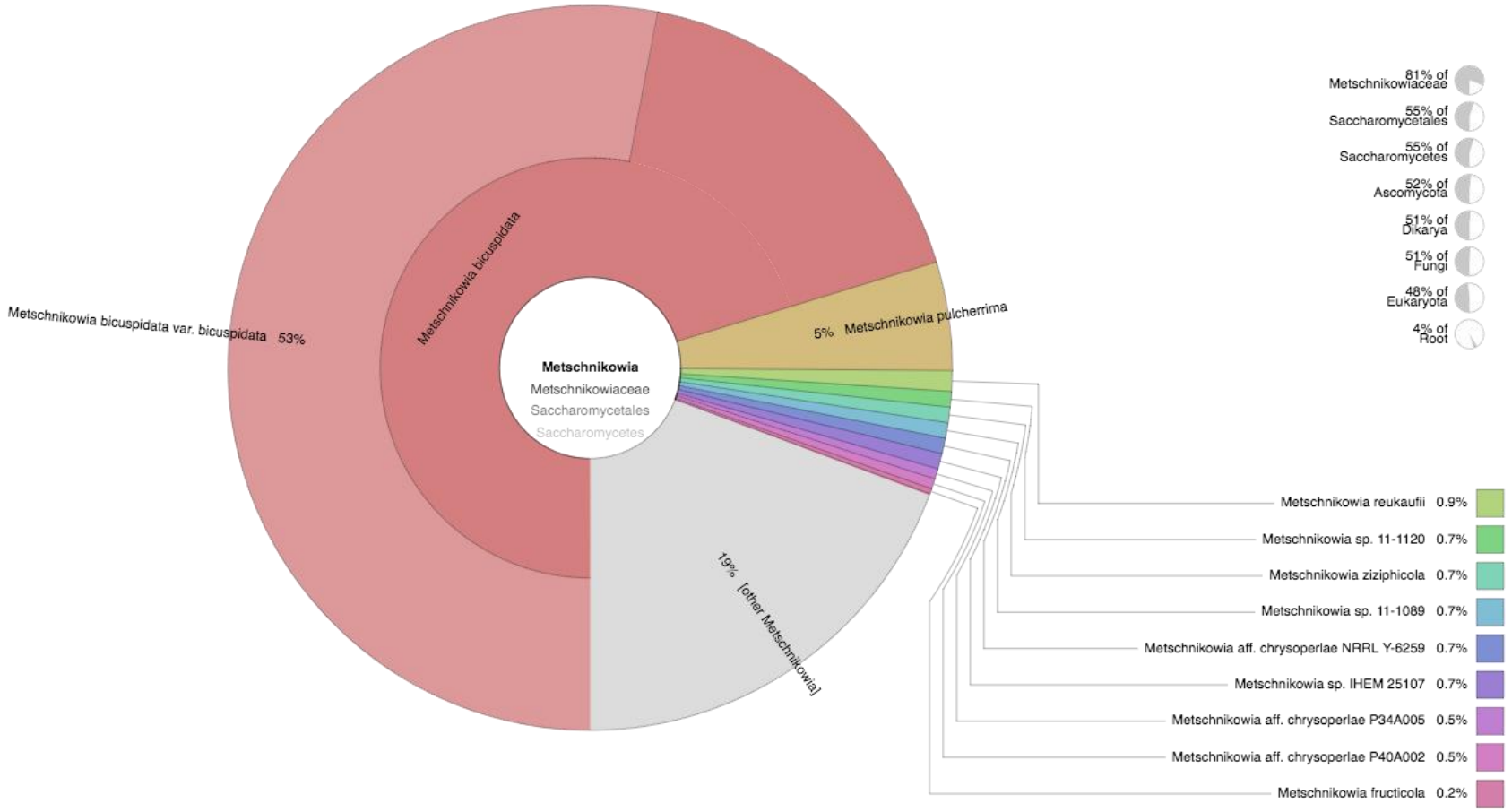


Figure 4.4: KRONA multi-layered pie chart showing the phylogeny relationship of the *M. pulcherrima* strain sequenced with other Metschnikowia genome/genetics data from online database.

4.3.3 DNA Assembly

In the event of the genomic sequencing of the strain WW002B, they were assembled using SPAdes version 3.11.0, the size is with the total length of 27,943,520 base pairs (27.9 Mbp). The largest contig being 379223 bp in size, a sum of 33,583,194 reads were run with 331.16 × coverage (and mean length of 287.82 bp), having the mapped read of 30,598,480 (91.1%). The contigs of 1000bp and larger are of 2179 count. The GC percentage is 45.77% which is a criterion for Ascomycetes as oppose to Basidiomycetes, which generally has a GC percentage of higher than 50% [152] .

Table 4.2: General characteristics of *M. pulcherrima* strain (WW02B) genome.

General features	**Contents**
Genome size	27943520 bp
Number of contigs	4217 bp
Contigs N50	25182 bp
Contigs N75	10750 bp
Largest contig	379223 bp
GC percentage	45.77 %
Protein-coding genes	4688 proteins

4.3.4 Protein coding

The proteins were aligned to *S. cerevisiae* (S288C) using BUSCO software application version 2 and showed a 90% positive hit, 4688 proteins were identified.

In another analysis using tRNAscan-SE, we identified 248 tRNA from the sequenced genome. From these list, 16 tRNAs were found with mismatch from the universal code. The $tRNA_{CAG}$ was repeatedly discovered to bind with Serine rather than Leucine as in table 4.1. This is a phenomena common in yeasts from the CUG-Ser clade, including *Candida albicans*, *C. cylindracea*, and many other Candidas (75 species); and also in *Pichia stipitis, Debaryomyces hansenii, Lodderomyces elongisporus* that use an altered genetic code, in which CUG codons are translated as Ser rather than Leu [153-155]. This is facilitated by the $tRNA_{CAG}^{Ser}$ that has the ability to be recognized by SerRS and LeuRS that proceeds to the synthesis of 2 different amonoacyl-tRNA: 1) $tRNA_{CAG}^{Ser}$; 2) $tRNA_{CAG}^{Leu}$ [155, 156].

Table 4.3: List of tRNA of *M. pulcherrima* strain (WW02B) with deviation of anti-codon from the universal genetic code.

	Sequence Name	tRNA Bounds		tRNA	Anti	Intron Bounds		Inf	Isotype	Isotype
		Begin	End	Type	Codon	Begin	End	Score	CM	Score
1	NODE_78_length_53821_cov_59.725798	35858	35765	Leu	CAG	35821	35810	52.7	Ser	85.5
2	NODE_84_length_52407_cov_56.123049	3417	3510	Leu	CAG	3454	3465	52.7	Ser	85.5
3	NODE_176_length_34465_cov_56.601171	1152	1245	Leu	CAG	1189	1200	52.7	Ser	85.5
4	NODE_415_length_18455_cov_58.281973	16579	16672	Leu	CAG	16616	16627	52.7	Ser	85.5
5	NODE_157_length_35920_cov_1661.375409	33815	33744	Arg	TCT	0	0	53.7	Gln	61
6	NODE_157_length_35920_cov_1661.375409	7938	7867	Met	CAT	0	0	59.6	Gly	57.2
7	NODE_157_length_35920_cov_1661.375409	12837	12767	Asn	GTT	0	0	57.8	Thr	56
8	NODE_157_length_35920_cov_1661.375409	8614	8544	Ile	GAT	0	0	59.8	Trp	52.8
9	NODE_157_length_35920_cov_1661.375409	10309	10239	Met	CAT	0	0	54.3	Thr	49.7
10	NODE_157_length_35920_cov_1661.375409	10554	10483	Arg	ACG	0	0	50.1	Thr	48.4
11	NODE_157_length_35920_cov_1661.375409	23175	23092	Ser	TGA	0	0	46.1	Leu	46.6
12	NODE_157_length_35920_cov_1661.375409	7780	7708	Glu	TTC	0	0	44.2	Gln	32.6
13	NODE_157_length_35920_cov_1661.375409	8013	7942	His	GTG	0	0	38.4	Cys	30.9
14	NODE_2_length_188721_cov_117.092718	52467	52397	Sup	TCA	0	0	30.8	His	29.7
15	NODE_157_length_35920_cov_1661.375409	27084	27013	Phe	GAA	0	0	38.7	His	29.3
16	NODE_157_length_35920_cov_1661.375409	10793	10722	Asp	GTC	0	0	33.5	Arg	27.2

From the listed deviations of anti-codons in table 4.3, other than the first four, the isotope scores of the rest are quite low to be take into account as significant.

In another analysis, we utilize the Venn chart (Figure 4.5) to observe the protein relationships between our strain of *M. pulcherrima*, *M. bicuspidata* var. *bicuspidata* (NRRL YB-4993) [Joint Genome Institution, U.S. Department of Energy], *Saccharomyces cerevisiae*, and *Yarraowia lipolytica*. *S. cerevisiae* is the well-studied yeast genome in the Saccharomyces Genome Database (SGD). *M. bicuspidata* is the only yeast from the Metschnikowia genus that has been fully sequenced. While *Y. lipolytica* is used because of its well established oleaginous property [157].

Table 4.3: Summary of the Protein quantity of all four species in the Venn chart.

Species	Proteins	Clusters	Singletons
Metschnikowia bicuspidata	5838	3803	1779
Saccharomyces cerevisiae	6692	3643	2102
Yarrowia lipolytica	6448	3756	1865
Metschnikowia pulcherrima	4688	2577	152

From the resulted Orthologous clusters Venn diagram, 2086 proteins were found to be shared between all four in this examination. *M. pulcherrima* and *M. bicuspidata* shared 156 proteins. While *M. pulcherrima* has 46 proteins uniquely of its own, *M. pulcherrima* and *Y. lipolytica*, shared only 5 similar proteins clusters. In the 5 protein clusters, 3 biological processes and 3 molecular functions are listed with their Hypergeometric test result of Gene Ontology (GO) enrichment has a p-value < 0.005 (Table 4.4).

The first two in the table, are involved in cell energy productions. Ketone bodies (acetoacetate, D-3-hydroxybutyrate, and acetone) can be used as an energy source as an alternative to glucose [158]. 3-oxoacid CoA-transferase, on the other hand, Catalyse the reaction between succinyl-CoA and a 3-oxo acid to produce succinate and a 3-oxo-acyl-CoA, which is a part of ketone bodies synthesis and degradation [158]. While the third in the list plays a part in ion transfer also in the production of energy. The shared genes of both species are very small in number to show any significance relationship in their ability to produce fatty acids.

Table 4.4: Hypergeometric test result of GO enrichment (p-value < 0.05)

GO ID	Name	p-value
GO:0046952	ketone body catabolic process	0.0024570024570024626
GO:0008260	3-oxoacid CoA-transferase activity	0.0024570024570024626
GO:0016712	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	0.0024570024570024626
GO:0033897	ribonuclease T2 activity	0.007354413940633074
GO:0006857	oligopeptide transport	0.009794845621172388
GO:0005215	transporter activity	0.04338838871790166

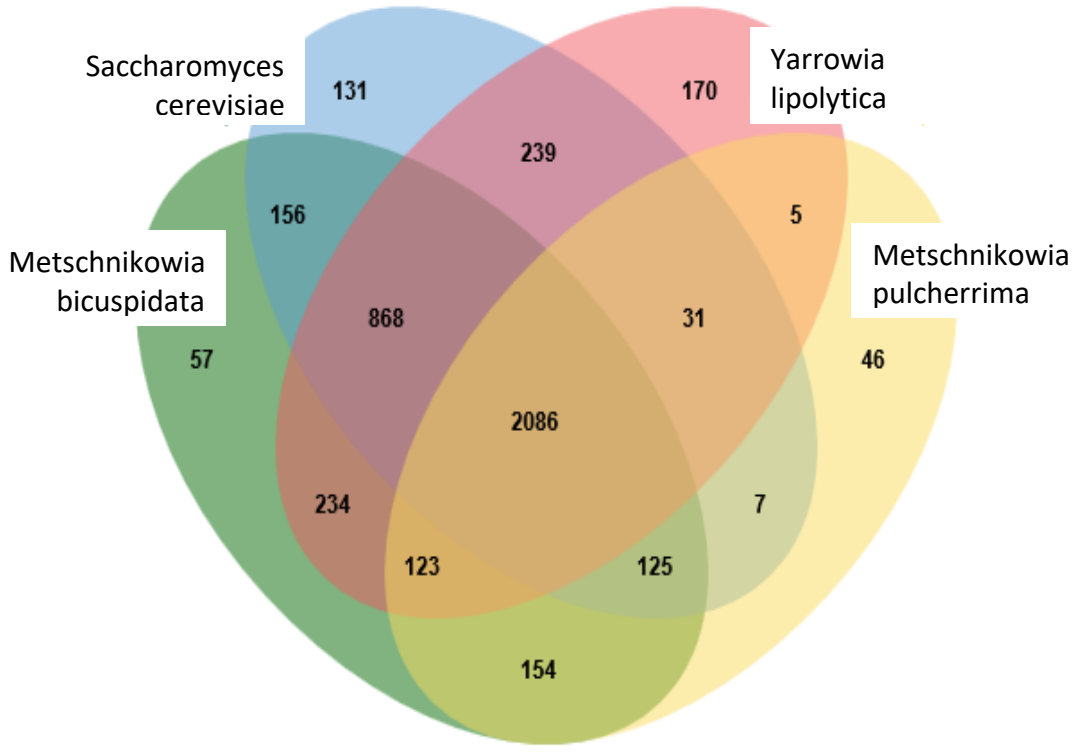


Figure 4.5: Venn diagramme representation of shared or unique genes in *M. pulcherrima* and comparison with those in *M. bicuspidata* var. *bicuspidata* (NRRL YB-4993), *S. cerevisiae*, and *Y. lipolytica.*

4.3 Conclusion

This is a de novo attempt to decipher the *M. pulcherrima* genome, from the Metschnikowia family that has not been genetically studied extensively. We know from this research project, that this species is an industrially beneficial species as it has been proven as a highly resilient strain with moderate capability to produce various long chain fatty acids, and some combinations of the product deemed constitute of palm oil. Studying the genome focusing on the metabolic pathway related to fatty acid metabolism, for example genes or protein involved in fatty acid biosynthesis; fatty acid elongation; fatty acid degradation; synthesis of degradation of ketone bodies; steroid biosynthesis; primary bile acid biosynthesis; glycerolipid metabolism; glycerophospholipid metabolism; ether lipid metabolism; arachidonic acid metabolism; linoleic acid metabolism; alpha-linolenic acid metabolism; sphigolipid metabolism; and biosynthesis of unsaturated fatty acid [159] is very much a crucial aspect to follow suit the genomic sequencing.

Only a few of the Metschnikowia family members has been put through the genomic sequencing, therefore, not much of comparison can be made. The ability of these species to adapt many types of harsh environments (such as high acidity of pH 4 to 5 & low temperature of 15°C to 20°C) where most of them are found, must make them possess some unique proteins to have such resilience. This is another group of genes that need to be discovered and even maybe biotechnologically enhanced. And perhaps a genotypic or phenotypic relationship between the ability to harness extreme conditions and lipid production may have some significance.

There are many more advantages with the deposit of genomics sequences in the Fungal world because fungal including yeasts are ideal organisms for comparative genomic studies in eukaryotes for the reason that they possess small and compact genomes and have a number of species that have been researched on thoroughly and continuously. The amount of information already available should be scrutinized to be beneficial for the understanding and even enhancements for these organisms.

Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this these a novel staggered Bioprospecting approach was developed and used to successfully isolate several resilient wild yeast strains, highly suited to industrial biotechnology applications. This method was designed to reduce an extensive phenotypic analyses of yeast strains from a given area, only isolating the strains with the best characteristics, thus cutting costs of culturing & identifying strains that would be of no industrial value. In this technique strains were selected that could grow in extremes of pH, presence of inhibitors and on multiple sugars. The technique was validated against a hardy strain of *Metschnikowia pulcherrima*.

Initially 12 strains were identified that could survive the conditions tested, and most of the hardy yeast strains from this collection (seven out of 12 identified) were from the *M. pulcherrima* group that includes *M. pulcherrima* itself and *Metschnikowia aff. chrysoperlae,* from identification using PCR-Sequencing technique of ITS region. Much of the ability of these strains to tolerate acidity, is due to their adaptation to grape skins, a highly acidic environment as well as the ability to produce, pulcherrimic acid, an antimicrobial pigment. *M. pulcherrima* was seen to bring down the acidity from pH 8 to as low as pH 4 in these experiments. pH tolerance is very useful in conditions where lignocellulose waste produced various acids as side products when they are being broken down, and in some instances using acid itself in the hydrolysis process as in acid hydrolysis.

In chapter 3, the 12 identified strains were cultured on a broader range of possible model hydrolysates composed to mimic various sources from around the world nd including a range of different inhibitors. The hydrolysates selected were wheat straw, corn stover, sugarcane bagasse, and palm kernel cake. All the strains could grow on these feedstocks under acidic conditions irrespective of the level of glucose in the samples, and there was little differentiation between species. Interestingly the ability to withstand inhibitors was directly correlated with pH, with high inhibitor concentrations only being able to be withstood at high ph. This is potentially due to neutralising the more acidic inhibitors aid the ability to withstand furfural and 5-HMF or that culturing under acidic conditions puts the organism under more strain, reducing the ability to oxidise the inhibitors to more benign compounds.

The strains were screened for lipid production, while none of the strains investigated produced high lipid, some did fall within the oleaginous range, with 20% or more of oil content from the biomass produced in a short 5-day growth. The lipid profile of the lipids produced were also examined. All the oils produced had simple lipid profiles akin to terrestrial crops with the major fatty acids palmitic acid, oleic acid and linoleic acid all present. The ratio of the lipids varied substantially and was also highly dependent on the temperature and inhibitor concentration. However, some strains were able to produce a palm oil substitute, with suitable levels of saturation. However, with the variation it is doubtful that a variable feedstock could be used to produce a lipid product with a fixed profile without some sort of genetic control.

In the final chapter, we deciphered the genomic sequence of one of the strains from our collection. We selected the *M. pulcherrima* strain (WW02B) to represent the whole study due to its capabilities being a resilient and robust strain and the need of more dependable biotechnological tools to be developed with the thorough knowledge of the genetic sequence of the species. With the yet to be successful genetic manipulations done to this species, it is deemed important that the genome information is harnessed and understood well for designing powerful tools for its genetic manipulations.

The addition of the genomic data on yeast is also beneficial in the genomic comparative studies that has been done to the already available data of fungal genomics. *M. pulcherrima* can be added to the pool of knowledge that has been accumulated since the first eukaryote, *S. cerevisiae* genome has been sequenced more than two decades ago.

5.2 Future work

In this work a novel bioprospecting method has been developed and applied to find organisms more suited to biotechnology.

Potentially, in any flora and fauna, hundreds of yeast strains from resilient genus such as *Metschnikowia* and *Candida* could be collected and tested for beneficial chemical production. Further bioprospecting of yeasts strain should be carried out from other less obvious origins. Other fruits, which are non-berries, for example, olives and tomatoes, are good potentials as they are known to have high levels of arabinose rather than glucose or xylose. Other that fruits and flowers, insects and their nymphs, for example bees & wasps, could also be considered as in many previous researches, insects are habitat for many yeast species – and the stomach areas are highly acidic.

Although high in resilience, the successful strains from the first experiment are still susceptible to concentrated inhibitors and extreme pH's. We can further test these strains in other harsh environmental conditions, such as lower temperatures, different sets of potent inhibitors and high pH's that are related to other types of industrial conditions pertaining hydrolysis processes. In addition, various carbon-sources that are not of lignocellulose nature must also be explored for the maximizing efforts in recycling for sustainability.

In addition, the lipid content of all of these strains is still too low, and while it is useful that some can produce lipid, this would need to be increased substantially to allow lipid accumulation to match current palm oil and industrial organisms. Many types of modification are possible to enhance this feature. The most obvious option would be genetic transformation of the strains into faster lipid-producing yeasts. In initiating a transformation experiment of a novel species, a number of steps could be taken:

1. Gene identification will be carried out for the specific phenotype.
2. A gene marker will be identified for the targeted gene. More information of the targeted gene from previous studies will be gathered to obtain the DNA sequence/genomic DNA.
3. Optimization of the DNA sequence must be carried out. In this case there is a number of parameters that can be explored as it is usually done in multi-objective. More research on the specific gene must be done before any focus can be made.

4. Selecting/creating suitable promoters is another aspect of optimization that must be considered.

Another angle that we can "develop" these newly studied strains is administering them through evolutionary processes and stresses. Mating, however pervasive in yeasts reproduction, has nonetheless, been a source of species evolution. In future experiments, we could selectively breed strains and subsequently, produce superior strains in lipid production. Other evolutionary studies that could come into play are exposing the strains to environmental changes or evolutionary stresses that could lead to high lipid production. Many new researches provide studies that gene expression can tune into physiological needs controlled by the environmental changes not only encourages organisms becoming more resilient but also fuel them into phenotypic variation and evolutionary innovation.

In this potential project, the portfolio of products targeted could be increased beyond palm oil substitute as many of these yeasts possess the ability to produce bioethanol for example. Other products that could be targeted could be acetic acid and other acids; cacao butter alternative, vitamins and amino acids producers.

In the final chapter, the seven strains of the *Metschnikowia pulcherrima group* were successfully grouped into 3-4 groups depending on the gene used. The inference statistical study that lead to these conclusions carries more information than just grouping of the strains. The analyses also carry the values of time, distance and linkage, hence, put some population and evolutionary values attached to them. From this data, we could also bring about evolutionary studies of these different strains. With additional strains and data incorporated, an extensive overview of this species could be detailed.

Aside from the *Metschnikowia pulcherrima group*, we also have other species in the collection. Two of *Candida*, one of each of lesser known oleaginous yeast genuses, *Rhodotorula, Meyerozyma* and *Wickerhamomyces*. These strains have shown striking similarity in resilience and oil production as the *Metschnikowia pulcherrima* group, hence, potentially relating them. *Meyerozyma guilliermondii* particularly has shown high growth rate in most of the experiments. Therefore, future work could be invested in attempting to further develop these yeast for industrial biotechnology.

Chapter 6

Material and Methods

6.1 Material and Methods

6.1.1 Materials and organisms

All chemicals were purchased from Sigma Aldrich unless otherwise stated, none of the chemicals used were purified prior to use. API® ID32 kit and API® NaCl 0.85 % Suspension Medium was bought from bioMérieux, Inc. DNA Sequencing method of PCR products was using the SeqPrimer tubes, and Primers purchased from Eurofin Genomics. Mix2Seq Sequencing Kit was purchased from eurofins Genomics. *M. pulcherrima* was obtained from the National Yeast Culture Collection (Norfolk, UK) and stored on sterile YPD agar plates at 4 °C. *M. pulcherrima* was re-plated every two months. The 12 strains of wild yeast strains were also stored under sterile conditions as above. A set of the strains were also kept in 25% Glycerine at -83°C for long term storage.

6.2 Materials and method for Chapter 2

6.2.1 *M. pulcherrima* maintenance in the lab

All *M. pulcherrima* lab strains are cultivated in yeast minimal medium (YMM)as adapted from Chatzifragkou et al. and Santamauro et al. (Table 6.1) before they were subjected to the bioprospecting culturing [160, 161]. The medium was prepared in deionised water in a Duran bottle, the pH was then adjusted to 5 using hydrochloric acid, then the calcium chloride added, then the Duran bottle was autoclaved at 121 °C, 20 min.

6.2.2 Minimal media for Stage 1

Minimal medium with additional lysine and antibiotics (Ampicilin & tetracycline) [**MML**] was composed of glucose 30 g/L, lysine 1.106 g/L, ampicillin 1 mL/L, tetracycline 1mL/L, yeasts extract 0.1 g/L, magnesium sulphate heptahydrate 1.5 g/L, zinc sulphate heptahydrate 0.02 g/L, manganese (II) sulphate monohydrate 0.06 g/L, iron(III) chloride hexahydrate 0.15 g/L, potassium phosphate monobasic 2.1 g/L, sodium phosphate dibasic 0.75 g/L, potassium sulphate 0.659 g/L and calcium chloride dihydrate 0.15 g/L in deionised water.

The acidic MML media **(MML_A)** was produced by addition of 0.375 g/L of tartaric acid and 2.117 g/L of potassium sodium tartrate tetrahydrate as buffer. The pH was then adjusted to pH 4.8. The basic MML media **(MML_B)** was produced by an additional 6.057 g/L of TRIS. The pH was adjusted to pH 9. All media was autoclaved prior to use.

Table 6.1: Yeast Minimal Medium for the maintenance of *M. pulcherrima* lab strains

Chemical	Chemical Formula	Quantity
Calcium chloride dehydrate	$CaCl_2.2H_20$	0.15 g/L
Magnesium sulphate heptahydrate	$MgSO_4.7H_2O$	1.5 g/L
Zinc sulphate heptahydrate	$ZnSO_4.7H_2O$	0.02 g/L
Manganese (II) sulphate monohydrate	$MnSO_4.H_2O$	0.06 g/L
Iron(III) chloride hexahydrate	$FeCl_3.6H_2O$	0.15 g/L
Potassium phosphate monobasic	KH_2PO_4	2.1 g/L
Sodium phosphate dibasic	Na_2PO_4	0.75 g/L
Sodium Sulphate	Na_2SO_4	0.537 g/L
Ammonium sulphate	$(NH_4)_2SO_4$	0.294 g/L
Ammonium chloride	NH_4Cl	0.572 g/L
Yeasts Extract		0.1 g/L
Glucose		30 g/L

6.2.3 Minimal media for Stage 2

Minimal medium with additional inhibitors (MMI) was composed of glucose 30 g/L, furfural (10mM) 0.8276 mL/L, Acetic Acid (60mM) 3.435 mL/L, formic Acid (60mM) 2.2266 mL/L, levullinic Acid (60mM) 6.1434 mL/L, 5-HMF (10mM) 1.1601 g/L, yeasts extract 0.1 g/L, magnesium sulphate heptahydrate 1.5 g/L, zinc sulphate heptahydrate 0.02 g/L, manganese (II) sulphate monohydrate 0.06 g/L, iron(III) chloride hexahydrate 0.15 g/L, potassium phosphate monobasic 2.1 g/L, sodium phosphate dibasic 0.75 g/L, potassium sulphate 0.659 g/L, ammonium sulphate 0.294 g/L, ammonium chloride 0.572 g/L and calcium chloride dihydrate 0.15 g/L in deionised water.

The acidic MML media **(MMI_A)** was produced by addition of 0.375 g/L of tartaric acid and 2.117 g/L of potassium sodium tartrate tetrahydrate as buffer. The pH was then adjusted to pH 4.8. The basic MML media **(MMI_B)** was produced by an additional 6.057 g/L of TRIS. The pH was adjusted to pH 9. All media was autoclaved prior to use.

6.2.4 Minimal media for Stage 3

Minimal media with xylose **(MMX**), was composed of xylose 30 g/L, yeasts extract 0.1 g/L, magnesium sulphate heptahydrate 1.5 g/L, zinc sulphate heptahydrate 0.02 g/L, manganese (II) sulphate monohydrate 0.06 g/L, iron(III) chloride hexahydrate 0.15 g/L,

potassium phosphate monobasic 2.1 g/L, sodium phosphate dibasic 0.75 g/L, potassium sulphate 0.659 g/L, ammonium sulphate 0.294 g/L, ammonium chloride 0.572 g/L and calcium chloride dihydrate 0.15 g/L in deionised water.

The acidic MML media **(MMX_A)** was produced by addition of 0.375 g/L of tartaric acid and 2.117 g/L of potassium sodium tartrate tetrahydrate as buffer. The pH was then adjusted to pH 4.8. The basic MML media **(MMX_B)** was produced by an additional 6.057 g/L of TRIS. The pH was adjusted to pH 9. All media was autoclaved prior to use.

6.2.5 Minimal media for Stage 4

Minimal media with arabinose and cellobiose **(MMAC**) was composed of arabinose 15 g/L, cellobiose 15g/L, yeasts extract 0.1 g/L, magnesium sulphate heptahydrate 1.5 g/L, zinc sulphate heptahydrate 0.02 g/L, manganese (II) sulfate monohydrate 0.06 g/L, iron(III) chloride hexahydrate 0.15 g/L, potassium phosphate monobasic 2.1 g/L, sodium phosphate dibasic 0.75 g/L, potassium sulphate 0.659 g/L, ammonium sulphate 0.294 g/L, ammonium chloride 0.572 g/L and calcium chloride dihydrate 0.15 g/L in deionised water.

The acidic MML media **($MMAC_A$)** was produced by addition of 0.375 g/L of tartaric acid and 2.117 g/L of potassium sodium tartrate tetrahydrate as buffer. The pH was then adjusted to pH 4.8. The basic MML media **($MMAC_B$)** was produced by an additional 6.057 g/L of TRIS. The pH was adjusted to pH 9. All media was autoclaved prior to use.

6.2.6 Wallerstein Laboratory (WL) Nutrient Agar

This was made up of yeast extract 4.0 g/L, tryptone 5.0 g/L, glucose 50.0 g/L, potassium dihydrogen phosphate 0.55 g/L, potassium chloride 0.425 g/L, calcium chloride 0.125 g/L, magnesium sulphate 0.125 g/L, ferric chloride 0.0025 g/L, manganese sulphate 0.125 g/L, bromocresol green 0.022, Agar 15.0, at pH 5.5 ± 0.2.

6.2.7 Inoculating and Culturing of *M. pulcherrima*

The yeast strain was maintained on yeast peptone dextrose (YPD) agar plates [10 g/L yeast extract, 20 g/L peptone, 20g/L glucose, 15 g/L agar] at 4 °C until used. The *M. pulcherrima* was grown in YPD media for 48 hours at 20°C and shaking at 180 rpm.

For the *M. pulcherrima* cultures 20 mL of the media, given above, was added to a 125-mL Erlenmeyer flask and innoculated with 2 mL of the *M. pulcherrima* inoculum in a

laminar flow cabinet. They were then incubated at 20°C and shaken at 180 rpm. This was done in three duplicates.

At 24 hour intervals, the OD at 600nm as well as the pH were recorded over the course of the cultures. All Erlenmeyer flasks were autoclaved at 121°C for 20 minutes prior to use.

The Optical Density (O.D) is measured using Light scattering techniques to monitor the concentration of the cultures. Light scattering is most closely related to the dry weight of the cells. Most visible light cannot penetrate a cell. When the light beam in a spectrophotometer hits a cell, the light is deflected from the light path, so some of the light never reaches the detector. The greater the number of cells in a sample, the more the light scattering that occurs. The light scattering ability of a cell depends on its size and geometry, so a calibration curve is necessary to extrapolate optical density measurements to cell number.

Light scattering is measured with the spectrophotometer set to report absorbance. Because the principles used to measure light scattering and absorbance are different, the amount of light scattered by a solution is referred to as its "optical density" rather than its "absorbance." The optical density of a sample analysed at 600 nm is abbreviated OD_{600}, with the subscript indicating the wavelength used for the measurement.

The spectrophotometer used in the lab was calibrated. Because the experiments run was taking the O.D. of unknown species, we only calibrated the machine to 0 (zero) with deionized water. As this is only an estimation of the O.D. to attain a surviving sample.

6.2.8 Sampling and preparation of the yeast strain collection

Bioprospecting sampling was carried out at the Mumford's Vineyard, Bath (21/10/2013). This vineyard is a 1.5-hectare plot overlooking the Avon valley east of Bath. In the vineyard various grapes are grown next to one another these include Kerner; Madeleine Angevine; Triomphe d'Alsace; Leon Millot; Reichensteiner.

The grapes and flowers were randomly and aseptically sampled directly from the vineyard (vineyard sample), one week prior to commercial harvesting. The grapes or

flowers were carefully picked, so as not to disturb the biota and packed into sterilised plastic falcon tubes prior to delivery to the lab.

In the lab, samples were rinsed in Ringer's solution. And the solutes are taken to be cultured directly to the 4-step filtering of yeasts.

6.2.8.1 Ringer's Solution

Sodium chloride 7.2 g, calcium chloride 0.17 g, potassium chloride 0.37 g. All reagents were dissolved into deionised water and the final volume was made to exactly 1 L. The pH was adjusted to pH 7.3-7.4. Once all the salts had thoroughly dissolved, the solution was sterilised by filtering through an ultrafiltration 0.22-µm membrane, the solution was aliquoted into single-use volumes (25-50 mL) and autoclaved.

6.2.7.2 Strains cultured in minimal media for Stage 1 (MML)

From each sample, 2 mL of the suspension was transferred into 20 mL of media in a 250-mL Erlenmeyer flasks of the MML, added in a laminar flow cabinet. They were then incubated at 20°C and shaken at 180 rpm. At 24 hour intervals, the OD (at 600nm) was recorded for 4 days. All Erlenmeyer flasks were autoclaved at 121°C for 30 minutes prior to use.

6.2.8.3 Strains cultured in minimal media for Stage 2 (MMI)

Surviving species from the earlier step were taken for this stage. To do this, 2 mL of the suspension was transferred into 20 mL of media in a 250-mL Erlenmeyer flasks of the MMI, added in a laminar flow cabinet. They were then incubated at 20°C and shaken at 180 rpm. At 24 hour intervals, the OD (at 600nm) was recorded for 7 days.

6.2.8.4 Strains cultured in minimal media for Stage 3 (MMX)

The surviving strainis from the earlier step was taken as a 2 mL suspension and transferred into 20 mL of media in a 250-mL Erlenmeyer flasks of the MML, added in a laminar flow cabinet. They were then incubated at 20°C and shaken at 180 rpm. At 24 hour intervals, the OD (at 600nm) was recorded for 5 days.

6.2.8.5 Strains cultured in minimal media for Stage 4 (MMAC)

Similarly, any surviving species derived from stage 3 was taken as a 2 mL of the suspension was transferred into 20 mL of media in a 250-mL Erlenmeyer flasks of the MML, added in a laminar flow cabinet. They were then incubated at 20°C and shaken at

180 rpm. At 24 hour intervals, the OD (at 600nm) was recorded for 4 days. At the end, the surviving yeast were taken on to be cultured on WL nutrient agar. Each agar plate had one or multiple species colonies. These were then separated onto individual YPD plates.

6.2.8.6 Wallerstein Laboratory (WL) Nutrient Agar

This was made up of yeast extract 4.0 g/L, tryptone 5.0 g/L, glucose 50.0 g/L, potassium dihydrogen phosphate 0.55 g/L, potassium chloride 0.425 g/L, calcium chloride 0.125 g/L, magnesium sulphate 0.125 g/L, ferric chloride 0.0025 g/L, manganese sulphate 0.125 g/L, bromocresol green 0.022, Agar 15.0, at pH 5.5 ± 0.2.

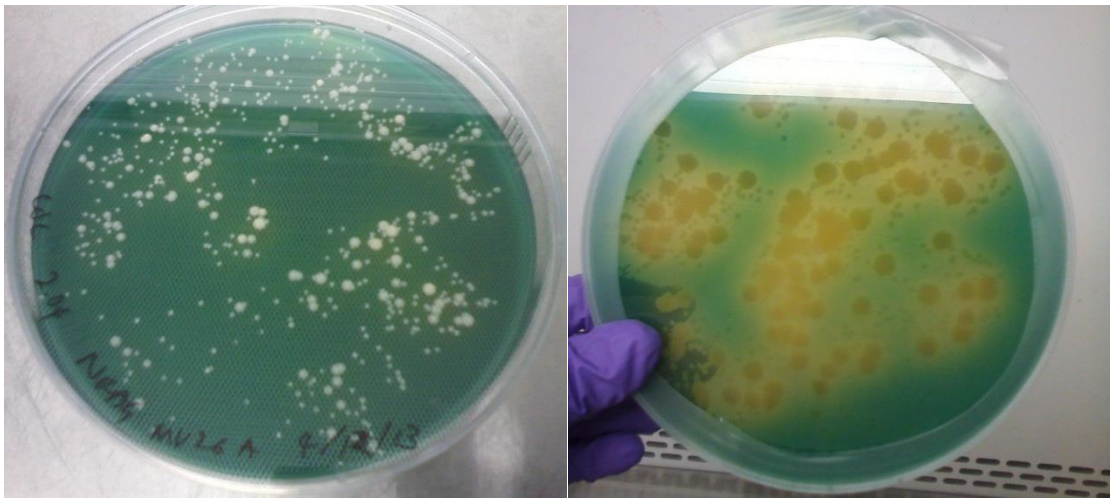


Figure 6.1: Unidentified surviving strains cultured on WL Nutrient agar

6.2.9 Identification of yeast strains

6.2.9.1 API® ID32 kit

One or several identical colonies were removed from the WL agar and transferred to YPD broth and cultured at 20 °C for 48 hours to promote growth. Then the growing sample was transferred onto a YDP agar plate to be grown for another 48 hours. A few colonies were transferred into the API® Suspension Medium (2 ml). The suspension turbidity was then adjusted to 2 McFarland. Next, approximately 250 µl of the suspension was transferred into API C Medium ampule. Immediately, the innoculated

ampule was homogenized, and 135 µl of the suspension was dispensed into each cupule of the API® ID32 kit.

The strip was then stored in an air-tight plastic container and incubated at 29°C ± 2°C for 24-48 hours. A humid atmosphere was created by filling a small amount of water in the container, which prevents the tests from drying out.

Figure 6.2: API® ID32 kit used for the initial collection identification of yeast strains

6.2.9.2 PCR products DNA sequencing

PCR from yeast colony: For direct PCR amplification, each reaction contained 13 µl of 2X concentration of DreamTaq Green PCR Master Mix (DreamTaq Green PCR Master Mix from Thermo Fisher Scientific contains DreamTaq™ DNA polymerase, optimized DreamTaq Green buffer, MgCl2 and dNTPs), 10 µl of autoclaved deionized water, 1 µl of each forward (ITS1: 5'-TCC GTA GGT GAA CCT GCG G-3') and reverse (LR3: GGT CCG TGT TTC AAG ACG G or ITS4: 5'-TCC TCC GCT TAT TGA TAT GC-3') primers (Table 4.1), to a total reaction volume of 25µl for each isolate. A small amount (approximately 1 mm^3) of a single colony was picked with a micropipette tip and added to the tubes as the DNA template.

DNA amplification was carried out in a PCR thermal cycler. The following PCR conditions were used to amplify ITS1-5.8SrDNA-ITS2 region:

> 94°C for 5 min followed by 25 cycles of 1 min (94°C); 1 min (55°C); 2 min (72°C) followed by a final extension of 72°C for 7 min.

Suitable positive and negative controls were included in each test throughout the experiments.

DNA extraction using About Xpedition™ Fungal/Bacterial DNA MiniPrep: For Strains that PCR was unsuccessful, DNA extraction using this system is applied. The steps were as follows:

1. 50-100 mg (wet weight) yeast cells that have been resuspended in 200 µl of water to a Xpedition BashingBead™ Lysis Tube. Then, 750 µl Xpedition™ Lysis/Stabilization Solution were added to the tube.
2. The suspended yeast was secured in a bead beater fitted with a 2 ml tube holder assembly and processed for a minimum 30 seconds.
3. The Xpedition BashingBead™ Lysis Tube was centrifuged in a microcentrifuge at 10,000 x g for 1 minute.
4. Afterwards, 400 µl supernatant to a Zymo-Spin™ IV Spin Filter (orange top) was transferred in a Collection Tube and centrifuged at 7,000 x g (~7,000 rpm) for 1 minute.
5. 1,200 µl of Fungal/Bacterial DNA Binding Buffer was added to the filtrate in the Collection Tube from Step 4.
6. Next, 800 µl of the mixture was transferred from Step 5 to a Zymo-Spin™ IIC Column in a Collection Tube and centrifuged at 10,000 x g for 1 minute.
7. The flow through from the Collection Tube was discard and Step 6 repeated.
8. 200 µl DNA Pre-Wash Buffer was added to the Zymo-Spin™ IIC Column in a new Collection Tube and centrifuge at 10,000 x g for 1 minute.
9. Then, 500 µl Yeast DNA Wash Buffer was added to the Zymo-Spin™ IIC Column and centrifuged at 10,000 x g for 1 minute.
10. Lastly, the Zymo-Spin™ IIC Column was transferred to a clean 1.5 ml microcentrifuge tube and 100 µl (25 µl minimum) DNA Elution Buffer was transferred directly to the column matrix. It is then centrifuged at 10,000 x g for 30 seconds to elute the DNA.

The DNA Was purified and the same PCR amplification technique is used to produce PCR products feasible for sequencing.

Detection of PCR products: 5 µl of amplicons was electrophoresed through a 1.5% agarose gel in TAE buffer (242g/L Tris base, 57.1m/L Glacial Acetic Acid, 18.6 g/L EDTA).

The PCR product are visualized using the ethidium bromide staining under UV irradiation.

DNA Sequencing: Mix2Seq from eurofins Genomics was used for fast sequencing of amplicons. The mixture was to be prepared in the Mix2Seq tubes as follows: 15 µl of autoclaved deionized water, 1 µl of primer (either ITS1-F or LR3-R (or ITS4-R), depending on which primer(s) the particular strain used in the PCR step), and 1 µl of PCR product to a total reaction volume of 17µl for each isolate.

Sequence Analysis: DNA sequences received from eurofins Genomics were then entered in to the BLAST (NCBI) portal for matches in the GenBank database. The results and BLAST accession numbers are shown in Table 2.7.

6.3 Materials and method for Chapter 3

6.3.1 Medium for the four sugar models with different inhibitor content

The same salt and nutrient concentrations were used as given above (table 6.1), the sugar contents were varied as given below. All specimens were grown in 96-well plates and the OD at 600nm were taken at end of 5-days growth.

6.3.1.1 Wheat Straw Hydrolyzate model

The sugar content was 12.80 g/L Xylose, 1.70 g/L Glucose, 2.60 g/L Arabinose [34].

6.3.1.2 Corn Stover Hydrolyzate model

The sugar content was 9.09 g/L Xylose, 2.13 g/L Glucose, 1.01 g/L Arabinose [124].

6.3.1.3 Sugarcane Bagasse Hydrolyzate model

The sugar content was 13.92 g/L Glucose, 7.123 g/L Xylose, 0.647 g/L Arabinose, 0.6414 g/L Glucuronic acid [162].

6.3.1.4 Palm Kernel Cake model

The sugar content was 2.31 g/L Glucose, 0.78 g/L Xylose, 0.33 g/L Arabinose, 0.57 g/L Galactose, 10.71 g/L Mannose [126]

6.3.1.5 Four concentrations of Inhibitors

The concentrations of inhibitors start with no inhibitors in the different sugar media. The inhibitor contents for low, medium and high loadings are given in tables 6.2-6.4.

Table 6.2: Inhibitors, at low concentration, most common in lignocellulose hydrolysate added to the culture medium [163]

Chemical	Formula	Quantity
Furfural (1 mmolar)	$C_5H_4O_2$	0.10 g/L
5-HMF (1 mmolar)	$C_6H_6O_3$	0.13 g/L
Acetic Acid (10 mmolar)	$C_2H_4O_2$	0.60 g/L
Formic Acid (10 mmolar)	CH_2O_2	0.46 g/L
Levullinic Acid (10 mmolar)	$C_5H_8O_3$	1.16 g/L

Table 6.3: Inhibitors, at medium concentration, most common in lignocellulose hydrolysate added to the culture medium [163].

Chemical	Formula	Quantity
Furfural (10 mmolar)	$C_5H_4O_2$	0.96 g/L
5-HMF (10 mmolar)	$C_6H_6O_3$	1.26 g/L
Acetic Acid (60 mmolar)	$C_2H_4O_2$	3.60 g/L
Formic Acid (60 mmolar)	CH_2O_2	2.76 g/L
Levullinic Acid (60 mmolar)	$C_5H_8O_3$	6.97 g/L

Table 6.4: Inhibitors, at high concentration, most common in lignocellulose hydrolysate added to the culture medium [163]

Chemical	Formula	Quantity
Furfural (60 mmolar)	$C_5H_4O_2$	5.76 g/L
5-HMF (60 mmolar)	$C_6H_6O_3$	7.57 g/L
Acetic Acid (200 mmolar)	$C_2H_4O_2$	12.01 g/L
Formic Acid (200 mmolar)	CH_2O_2	9.21 g/L
Levullinic Acid (200 mmolar)	$C_5H_8O_3$	23.22 g/L

6.3.1.6 A range of 9 different pHs.

The media were made up of the four different sugar models and with four levels of inhibitors. Then, each of the 16 resulting media were adjusted to 9 different pHs. The range examined were pH 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, and 7.0. This was done with adding 4.503 g/L of Tartaric acid (which is equivalent to 30mM) and 2.117 g/L of Potassium sodium tartrate tetrahydrate as buffers. Later each mixture is adjusted to the required pHs by adding an appropriate amount of hydrochloric acid.

6.3.2 Lipid assessment

The same media was used as above except (section 6.3.2) except with a 2:1 ratio of the sugars glucose & xylose (a total of 30 g/L). Two different pHs (pH 5 and 6), and inhibitors (low and medium) mixture as given below. The cultures were grown in 125-mL Erlenmeyer flasks shaken at 180 rpm in an incubator of 20°C over 7 days.

6.3.3.5 Cell dry mass

The cell dry mass was materialized by centrifuging the fermentation broth at 6000 rpm, decanting the supernatant, freeze or oven drying (40 °C) the pellet for 24-48 hours then weighing the yeast biomass. Where optical density (OD) used to estimate cell dry mass, a range of optical densities and a calibration curve calculated.6.3.3.6 Oil extraction

The standard 'Bligh and Dyer' method [91] was used to extract oils and then was modified based on the procedure of Pan *et al* [164]. Biomass from above were freeze-dried, then heated at 50 °C in 4M HCL for 60 min, then cooled, and 1:1 chloroform: methanol added and stirred overnight. The chloroform layer solvent removed using the rota-evaporator under vacuum condition. The extracted oil was weighed, and the process repeated until there was no change in the mass of oil.

6.3.3 FAME Analysis

The oil yielded from the extraction above was first dissolved in n-hexane. The resulting lipid was transesterified using 99% methanol and 1% sulphuric acid, heated in Ace Pressure Tubes at 60°C for 24 hours. After it is cooled down, the solvent (hexane) layer was carefully removed by pipette into tared vials. The FAME was washed with water three times according to a standard literature procedure [165]. The resulting esters were quantified by GC-MS with the resulting FAME component compared to known

FAME standards. GC-MS analysis was carried out using an Agilent 7890A Gas Chromatograph equipped with a capillary column (60m × 0.250mm internal diameter) coated with DB-23 (50%-cyanpropyl]-methylpolysiloxne) stationary phase (0.25µm film thickness) and a He mobile phase (flow rate: 1.2ml/min) coupled with an Agilent 5975C inert MSD with Triple Axis Detector. FAME samples were initially dissolved in 2ml of hexane and 1µl of this solution was loaded onto the column, pre-heated to 150°C. This temperature was held for 5 mins and then heated to 250°C at a rate of 4°C/min and then held for 2 mins.

6.4 Materials and method for Chapter 4

6.4.1 Culturing the yeast strain for the accumulation of biomass

Strain No. 5 was selected was selected for this protocol. About 10 mL of the culture grown in Yeast Malt Extract agar (malt extract, 30 g/L , mycological peptone, 5 g/L, agar, 15 g/L) plate were transferred into 2 1 L Erlenmeyer flasks. They were left in a 20˚C shaking Incubator for 3 days in SMB broth.

At the end of day 3, a very thick culture was observed and the broth was collected in 50 mL sterile conical tubes. They were all then centrifuged at 3000 rpm for 15 minutes. The supernatant were discarded & the pellets were used in the next steps. The rest of the unused centrifuged biomass was then freeze at -21˚C.

6.4.2 DNA extraction of strain No. 5 (*Metschnikowia pulcherrima*)

About 2 mL of biomass were re-suspended from the culture above by pipette, mixing in 100 µl of sterile Phosphate Buffer solution (PBS). 10 mL of Lysis Buffer [10 mM Tris-Cl, pH 8.0, 25 mM EDTA, pH 8.0, 0.5% (w/v) SDS, 20 µg/ml RNase A (added fresh before use)] was added and then the mixture was incubated for 1 hour at 37°C [166].

For cleaning up, 0.4% Polyvinylpyrrolidone (PVP) and 0.3% β-mercaptoethanol are added as reducing agent and antioxidant agent [167].

The next step is to add 50 µl Proteinase K (from stock concentration of 20 mg/mL). The mixture is slowly rotated end-over-end 10 times. It is then incubated at 50°C for 3 hours mixing by rotating end-over-end 10 times after 1 and 2 hours [166].

At this point, an equal volume of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) was added into the mixture. It was vortex for a few seconds to mix well and form an emulsion. To separate the emulsions, it is then centrifuged at 12000-14000 rpm for 5min. Three layers of DNA suspension, interphase and phenol formed. The top aqueous phase of DNA was pipetted off while carefully avoiding taking in any precipitated material from the interphase or any phenol [168].

Then Chloroform Back Extraction is applied by adding equal volumes of the chloroform/isoamyl alcohol solution produced earlier. The tube is vortexed vigorously for 1 minute. And then spun at a high speed (of 1500 rpm) for 5 minutes. As much of the top aqueous solution is removed and place into a new tube. Again, evading picking up any of the chloroform/isoamyl alcohol phase (adapted from PacBio SampleNet).

For cleaning up of Ethanol, we use Ammonium Acetate, to make a final concentration of 0.75 M. The solution was mixed well. 2.5X volume of 100% ethanol was then added and mixed well and incubated at -20 °C for 1 hour. It was later centrifuged for 20 minutes in a 4 °C centrifuge at 3000 rpm. Next, the supernatant is carefully decanted without disturbing the DNA (pellet). The Ethanol wash steps were then repeated, this time with 80% Ethanol. A quick spin of the supernatant on table top centrifuge was carried out to draw residual Ethanol to the bottom. Residual Ethanol was removed with a P20 pipette carefully, not to disturb the pellet. The transparent pallets are then air dried for 1-2 minutes. They are then resuspended in appropriate volume Elution Buffer.

The DNa quality were then examined using Qubit® 2.0 fluorometer reading (with QUANT-iT dsDNA assay) and NanoDrop® 1000 UV-Vis Spectrophotometer.

6.4.3 Library Preparation for DNA sequencing

The samples were prepared for sequencing using Nextflex Rapid DNA protocol. The general steps are as follows:

1. End-Repair & Adenylation
2. Adapter Ligation
3. PCR Amplification
4. Bead Size Selection

The samples were then accessed using Qubit® flourometer.

6.4.4 Genomic DNA Sequencing

The sequencing libraries were run on a Hiseq 2500 in rapid run mode yielding 16.8 million 300 base pair paired end reads.

6.4.4 Protein coding

Annotation was performed using Maker version 2.31.9 which employs AUGUSTUS version 3.2.3 as gene prediction tool.

The data was also BLASTx against uniprot database for Protein Homology evidence.

BUSCO software application version 2 was used to decipher the protein

6.4.5 Other analysist

tRNA Scan was done on using the tRNAscan-SE application software

Using OrthoVenn for comparison and annotation of orthologous gene clusters among multiple species. Orthologous clusters Venn diagram was produced to compare between *M. pulcherrima*, *M. bicuspidata* var. *bicuspidata*, *Saccharomyces cerevisiae*, and *Yarraowia lipolytica*.

***6.4.3, 6.4.4, and 6.4.5 were carried out by the technical team at Exeter Sequence Service.**

Chapter 7

References

References

1. Thomas, A., *Fats and Fatty Oils*, in *Ullmann's Encyclopedia of Industrial Chemistry*. 2000, Wiley-VCH Verlag GmbH & Co. KGaA.
2. OCED, F., *OECD-FAO agricultural outlook 2011-2020*. 2011, Paris, France: Organisation for Economic Co-operation and Development (OECD) Publishing. doi.
3. MPOB. *Official Palm Oil Information Source*. Available from: http://www.palmoilworld.org/.
4. Adib, R., *Renewables 2015 Global Status Report.* 2015.
5. Lai, O.-M., *Palm Oil Production, Processing, Characterization, and Uses*. Palm Oil, ed. C.-P. Tan and C.C. Akoh. 2012, Burlington: Burlington : Elsevier Science.
6. Lai, O.-M., C.-P. Tan, and C.C. Akoh, *Palm Oil - Production, Processing, Characterization, and Uses*. AOCS Press.
7. Corley, R.H.V. and P. Tinker, *The oil palm*. 2008: John Wiley & Sons.
8. Berger, K.G., *Palm Oil Products - Why and How to Use them.* Fette, Seifen, Anstrichmittel, 1986. **88**(7): p. 250-258.
9. Broich, M., et al., *Remotely sensed forest cover loss shows high spatial and temporal variation across Sumatera and Kalimantan, Indonesia 2000-2008.* Environmental Research Letters, 2011. **6**(1).
10. Margono, B.A., et al., *Mapping and monitoring deforestation and forest degradation in Sumatra (Indonesia) using Landsat time series data sets from 1990 to 2010.* Environmental Research Letters, 2012. **7**(3).
11. Tauli-Corpuz, V. and P. Tamang, *Oil Palm and Other Commercial Tree Plantations, Monocropping: Impacts on Indigenous Peoples' Land Tenure and Resource Management Systems and Livelihoods.*, P.F.o.I. Issues, Editor. 2007, United Nations: New York.
12. Hinrichsen, N., *Commercially available alternatives to palm oil.* Lipid Technology, 2016. **28**(3-4): p. 65-67.
13. Godfray, H.C.J., et al., *Food security: the challenge of feeding 9 billion people.* science, 2010. **327**(5967): p. 812-818.

14. Escobar, J.C., et al., *Biofuels: Environment, technology and food security.* Renewable & Sustainable Energy Reviews, 2009. **13**(6-7): p. 1275-1287.
15. Clark, S.J.W., L. Schrock, M.D. Piennaar, P.G. , *Methyl and ethyl soybean esters as renewable fuels for diesel engines.* JAOCS, 1984. **61**: p. 1632–8.
16. Fazal, M., A. Haseeb, and H. Masjuki, *Biodiesel feasibility study: an evaluation of material compatibility; performance; emission and engine durability.* Renewable and Sustainable Energy Reviews, 2011. **15**(2): p. 1314-1324.
17. Monyem, A. and J. H. Van Gerpen, *The effect of biodiesel oxidation on engine performance and emissions.* Biomass and Bioenergy, 2001. **20**(4): p. 317-325.
18. Tat, M.E. and J.H. Van Gerpen, *The kinematic viscosity of biodiesel and its blends with diesel fuel.* Journal of the American Oil Chemists' Society, 1999. **76**(12): p. 1511-1513.
19. Laursen, W., *Students take a green initiative.* Tce, 2006(774-5): p. 32-34.
20. Mohr, A. and S. Raman, *Lessons from first generation biofuels and implications for the sustainability appraisal of second generation biofuels.* Energy Policy, 2013. **63**: p. 114-122.
21. Bozbas, K., *Biodiesel as an alternative motor fuel: Production and policies in the European Union.* Renewable and Sustainable Energy Reviews, 2008. **12**(2): p. 542-552.
22. Lu, J., C. Sheahan, and P. Fu, *Metabolic engineering of algae for fourth generation biofuels production.* Energy & Environmental Science, 2011. **4**(7): p. 2451-2466.
23. Kulkarni, M.G. and A.K. Dalai, *Waste cooking oil an economical source for biodiesel: a review.* Industrial & engineering chemistry research, 2006. **45**(9): p. 2901-2913.
24. Jenkins, R.W., et al., *Effect of the Type of Bean, Processing, and Geographical Location on the Biodiesel Produced from Waste Coffee Grounds.* Energy & Fuels, 2014. **28**(2): p. 1166-1174.
25. Atabani, A.E., et al., *A comprehensive review on biodiesel as an alternative energy resource and its characteristics.* Renewable & Sustainable Energy Reviews, 2012. **16**(4): p. 2070-2093.

26. Chuck, C.J., et al., *Liquid transport fuels from microbial yeasts – current and future perspectives.* Biofuels, 2014. **5**(3): p. 293-311.
27. Sainz, M.B., *Commercial Cellulosic Ethanol The Role of Plant-Expressed Enzymes*. Biofuels: Global Impact on Renewable Energy, Production Agriculture, and Technological Advancements, ed. D. Tomes, P. Lakshmanan, and D. Songstad. 2011, New York: Springer. 237-264.
28. Carriquiry, M.A., X.D. Du, and G.R. Timilsina, *Second generation biofuels: Economics and policies.* Energy Policy, 2011. **39**(7): p. 4222-4234.
29. Riffell, S., et al., *Potential Biodiversity Response to Intercropping Herbaceous Biomass Crops on Forest Lands.* Journal of Forestry, 2012. **110**(1): p. 42-47.
30. OCED, F., *OECD-FAO Agricultural Outlook 2005-2014*. 2005, Paris, France: Organisation for Economic Co-operation and Development (OECD) Publishing. .
31. FAO, R., *FAOSTAT database.* Food and Agriculture Organization of the United Nations, Rome, Italy, 2013.
32. Somerville, C., *The billion-ton biofuels vision.* Science, 2006. **312**(5778): p. 1277-1277.
33. Nigam, P.S. and A. Singh, *Production of liquid biofuels from renewable resources.* Progress in Energy and Combustion Science, 2011. **37**(1): p. 52-68.
34. Kumar, P., et al., *Methods for Pretreatment of Lignocellulosic Biomass for Efficient Hydrolysis and Biofuel Production.* Industrial & Engineering Chemistry Research, 2009. **48**(8): p. 3713-3729.
35. Pirozzi, D., et al., *Exploitation of agricultural biomasses to produce iigeneration biodiesel*, in *Chemical Engineering Transactions*. 2013. p. 175-180.
36. Roy, P., et al., *A Review of Life Cycle Assessment (LCA) of Bioethanol from Lignocellulosic Biomass.* Jarq-Japan Agricultural Research Quarterly, 2012. **46**(1): p. 41-57.
37. Sarkar, N., et al., *Bioethanol production from agricultural wastes: An overview.* Renewable Energy, 2012. **37**(1): p. 19-27.

38. Malherbe, S. and T.E. Cloete, *Lignocellulose biodegradation: fundamentals and applications.* Reviews in Environmental Science and Biotechnology, 2002. **1**(2): p. 105-114.

39. Perlack, R.D., et al., *US billion-ton update: biomass supply for a bioenergy and bioproducts industry.* 2011.

40. WASDE, U., *United States Department of Agriculture Foreign Agricultural Service.* Production, Supply and Distribution Online. Disponível em: www. fas. usda. gov/psd, 2006.

41. Thornley, P. *Clean Energy from Rice Straw*. 2012 [cited 2015 09/07/2015]; Available from: http://www.supergen-bioenergy.net/research-projects/rice-straw-project/.

42. Millati, R., et al., *Biological pretreatment of lignocelluloses with white-rot fungi and its applications: a review.* BioResources, 2011. **6**(4): p. 5224-5259.

43. Hu, G., J.A. Heitmann, and O.J. Rojas, *Feedstock pretreatment strategies for producing ethanol from wood, bark, and forest residues.* BioResources, 2008. **3**(1): p. 270-294.

44. Iqbal, H.M.N., G. Kyazze, and T. Keshavarz, *Advances in the valorization of lignocellulosic materials by biotechnology: An overview.* BioResources, 2013. **8**(2): p. 3157-3176.

45. *Ethanol from Lignocellulose*, in *Biofuels and Bioenergy*. 2012, CRC Press. p. 93-146.

46. Grabber, J.H., *How do lignin composition, structure, and cross-linking affect degradability? A review of cell wall model studies.* Crop Science, 2005. **45**(3): p. 820-831.

47. Ralph, J., et al., *Lignins: Natural polymers from oxidative coupling of 4-hydroxyphenyl- propanoids.* Phytochemistry Reviews, 2004. **3**(1): p. 29-60.

48. Scheller, H.V. and P. Ulvskov, *Hemicelluloses.* Annu Rev Plant Biol, 2010. **61**: p. 263-89.

49. Kuhad, R.C., A. Singh, and K.-E.L. Eriksson, *Microorganisms and enzymes involved in the degradation of plant fiber cell walls*, in *Biotechnology in the pulp and paper industry*. 1997, Springer. p. 45-125.

50. McKendry, P., *Energy production from biomass (part 1): overview of biomass.* Bioresource Technology, 2002. **83**(1): p. 37-46.

51. Bilal, M., et al., *Biotransformation of lignocellulosic materials into value-added products—A review.* International Journal of Biological Macromolecules, 2017. **98**: p. 447-458.

52. Sathitsuksanoh, N., A. George, and Y.H.P. Zhang, *New lignocellulose pretreatments using cellulose solvents: a review.* Journal of Chemical Technology and Biotechnology, 2013. **88**(2): p. 169-180.

53. Subhedar, P.B. and P.R. Gogate, *Intensification of Enzymatic Hydrolysis of Lignocellulose Using Ultrasound for Efficient Bioethanol Production: A Review.* Industrial & Engineering Chemistry Research, 2013. **52**(34): p. 11816-11828.

54. Limayem, A. and S.C. Ricke, *Lignocellulosic biomass for bioethanol production: Current perspectives, potential issues and future prospects.* Progress in Energy and Combustion Science, 2012. **38**(4): p. 449-467.

55. Cullis, I.F. and S.D. Mansfield, *Optimized Delignification of Wood-Derived Lignocellulosics for Improved Enzymatic Hydrolysis.* Biotechnology and Bioengineering, 2010. **106**(6): p. 884-893.

56. Wong, K.K.Y., et al., *THE RELATIONSHIP BETWEEN FIBER POROSITY AND CELLULOSE DIGESTIBILITY IN STEAM-EXPLODED PINUS-RADIATA.* Biotechnology and Bioengineering, 1988. **31**(5): p. 447-456.

57. Mood, S.H., et al., *Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment.* Renewable & Sustainable Energy Reviews, 2013. **27**: p. 77-93.

58. Sun, Y. and J. Cheng, *Hydrolysis of lignocellulosic materials for ethanol production: a review.* Bioresource Technology, 2002. **83**(1): p. 1-11.

59. Saritha, M., A. Arora, and Lata, *Biological Pretreatment of Lignocellulosic Substrates for Enhanced Delignification and Enzymatic Digestibility.* Indian Journal of Microbiology, 2012. **52**(2): p. 122-130.

60. Ravikumar, R., et al., *Innovative and intensified technology for the biological pretreatment of agro waste for ethanol production.* Korean Journal of Chemical Engineering, 2013. **30**(5): p. 1051-1057.
61. Gírio, F.M., et al., *Hemicelluloses for fuel ethanol: A review.* Bioresource Technology, 2010. **101**(13): p. 4775-4800.
62. Jonsson, L.J., B. Alriksson, and N.O. Nilvebrant, *Bioconversion of lignocellulose: inhibitors and detoxification.* Biotechnology for Biofuels, 2013. **6**.
63. Cragg, S.M., et al., *Lignocellulose degradation mechanisms across the Tree of Life.* Current Opinion in Chemical Biology, 2015. **29**: p. 108-119.
64. Palmqvist, E. and B. Hahn-Hagerdal, *Fermentation of lignocellulosic hydrolysates. I: inhibition and detoxification.* Bioresource Technology, 2000. **74**(1): p. 17-24.
65. Palmqvist, E. and B. Hahn-Hagerdal, *Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition.* Bioresource Technology, 2000. **74**(1): p. 25-33.
66. Imai, T. and T. Ohno, *THE RELATIONSHIP BETWEEN VIABILITY AND INTRACELLULAR PH IN THE YEAST SACCHAROMYCES-CEREVISIAE.* Applied and Environmental Microbiology, 1995. **61**(10): p. 3604-3608.
67. Nilvebrant, N.O., et al., *Identification of fermentation inhibiting compounds from acid hydrolysis of wood*

ISWPC - 9TH INTERNATIONAL SYMPOSIUM ON WOOD AND PULPING CHEMISTRY - POSTER PRESENTATIONS. 1997: p. 791-794.

68. Palmqvist, E., J.S. Almeida, and B. Hahn-Hagerdal, *Influence of furfural on anaerobic glycolytic kinetics of Saccharomyces cerevisiae in batch culture.* Biotechnology and Bioengineering, 1999. **62**(4): p. 447-454.
69. Ratledge, C. and J.P. Wynn, *The biochemistry and molecular biology of lipid accumulation in oleaginous microorganisms*, in *Advances in Applied Microbiology*. 2002, Academic Press. p. 1-51.
70. Papanikolaou, S. and G. Aggelis, *Lipids of oleaginous yeasts. Part II: Technology and potential applications.* European Journal of Lipid Science and Technology, 2011. **113**(8): p. 1052-1073.

71. Hu, C., et al., *Simultaneous utilization of glucose and xylose for lipid production by Trichosporon cutaneum.* Biotechnology for Biofuels, 2011. **4**.
72. Ratledge, C. and Z. Cohen, *Microbial and algal oils: Do they have a future for biodiesel or as commodity oils?* Lipid Technology, 2008. **20**(7): p. 155-160.
73. Santomauro, F., et al., *Low-cost lipid production by oleaginous yeast cultured in non-sterile conditions using waste resources.* Biotechnology for Biofuels, 2014. **7**(1): p. 34-44.
74. Firn, R.D., *Bioprospecting–why is it so unrewarding?* Biodiversity & Conservation, 2003. **12**(2): p. 207-216.
75. *Bioprospecting. Encyclopeadia. http://www.enviropaedia.com/.*
76. Acharya, S. and A. Chaudhary, *Bioprospecting thermophiles for cellulase production: a review.* Brazilian Journal of Microbiology, 2012. **43**: p. 844-856.
77. Headon, D. and G. Walsh, *The industrial production of enzymes.* Biotechnology advances, 1994. **12**(4): p. 635-646.
78. Mutanda, T., et al., *Bioprospecting for hyper-lipid producing microalgal strains for sustainable biofuel production.* Bioresource Technology, 2011. **102**(1): p. 57-70.
79. Mateo, N., W. Nader, and G. Tamayo, *Bioprospecting.* Encyclopedia of biodiversity, 2001. **1**: p. 471-488.
80. Chin, Y.-W., et al., *Drug discovery from natural sources.* The AAPS Journal, 2006. **8**(2): p. E239-E253.
81. Wynberg, R., *Hot air over Hoodia.* Seedling, GRAIN, 2010. **13**.
82. Walsh, G., *Biopharmaceutical benchmarks: Less than 20 years after the first biopharmaceutical, insulin, was approved, a new generation of products is superseding replacement proteins and enzyme therapies in biotechnology's pipeline.* Nature Biotechnology, 2000. **18**(8): p. 831-833.
83. Reid, W.V., *The economic realities of biodiversity.* Issues in Science and Technology, 1993. **10**(2): p. 48-55.

84. Tamayo, G., W. Nader, and A. Sittenfeld, *Biodiversity for bioindustries.* BIOTECHNOLOGY IN AGRICULTURE SERIES, 1997: p. 255-280.

85. Witt, W., et al., *Thrombolytic properties of Desmodus rotundus (vampire bat) salivary plasminogen activator in experimental pulmonary embolism in rats.* Blood, 1992. **79**(5): p. 1213-1217.

86. Borroni-Bird, C.E. *Smarter vehicles*. in *Smart Structures and Materials' 97*. 1997. International Society for Optics and Photonics.

87. Rausser, G.C. and A.A. Small, *Valuing Research leads: Bioprospecting and the Conservation of Genetic Resources.* Journal of Political Economy, 2000. **108**(1): p. 173-206.

88. Dhillion, S.S., et al., *Bioprospecting: Effects on Environment and Development.* AMBIO: A Journal of the Human Environment, 2002. **31**(6): p. 491-493.

89. Artuso, A., *Bioprospecting, Benefit Sharing, and Biotechnological Capacity Building.* World Development, 2002. **30**(8): p. 1355-1368.

90. Khan, S.A., et al., *Prospects of biodiesel production from microalgae in India.* Renewable and Sustainable Energy Reviews, 2009. **13**(9): p. 2361-2372.

91. Bligh, E.G. and W.J. Dyer, *A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION.* Canadian Journal of Biochemistry and Physiology, 1959. **37**(8): p. 911-917.

92. Spolaore, P., et al., *Commercial applications of microalgae.* Journal of bioscience and bioengineering, 2006. **101**(2): p. 87-96.

93. Lee, Y.-J., et al., *Screening Wild Yeast Strains for Alcohol Fermentation from Various Fruits.* Mycobiology, 2011. **39**(1): p. 33-39.

94. Choudhary, J., S. Singh, and L. Nain, *Bioprospecting thermotolerant ethanologenic yeasts for simultaneous saccharification and fermentation from diverse environments.* Journal of Bioscience and Bioengineering, 2017. **123**(3): p. 342-346.

95. Duarte, S.H., et al., *Exploration of Brazilian biodiversity and selection of a new oleaginous yeast strain cultivated in raw glycerol.* Bioresource Technology, 2013. **138**: p. 377-381.

96. Chen, X., et al., *Screening of Oleaginous Yeast Strains Tolerant to Lignocellulose Degradation Compounds.* Applied Biochemistry and Biotechnology, 2009. **159**(3): p. 591.

97. Nelkin, D., *Molecular metaphors: the gene in popular discourse.* Nat Rev Genet, 2001. **2**(7): p. 555-559.

98. Ortholand, J.Y. and A. Ganesan, *Natural products and combinatorial chemistry: back to the future.* Curr Opin Chem Biol, 2004. **8**(3): p. 271-80.

99. Zotchev, S.B., O.N. Sekurova, and L. Katz, *Genome-based bioprospecting of microbes for new therapeutics.* Current Opinion in Biotechnology, 2012. **23**(6): p. 941-947.

100. Bentley, S.D., et al., *Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2).* Nature, 2002. **417**(6885): p. 141-147.

101. Zotchev, S.B., *Marine actinomycetes as an emerging resource for the drug development pipelines.* Journal of Biotechnology, 2012. **158**(4): p. 168-175.

102. Chater, K.F. and M.J. Bibb, *Regulation of bacterial antibiotic production.* Biotechnology Set, Second Edition, 1997: p. 57-105.

103. Zazopoulos, E., et al., *A genomics-guided approach for discovering and expressing cryptic metabolic pathways.* Nat Biotech, 2003. **21**(2): p. 187-190.

104. Medema, M.H., et al., *antiSMASH: rapid identification, annotation and analysis of secondary metabolite biosynthesis gene clusters in bacterial and fungal genome sequences.* Nucleic acids research, 2011. **39**(suppl 2): p. W339-W346.

105. Velásquez, J.E. and W.A. Van der Donk, *Genome mining for ribosomally synthesized natural products.* Current opinion in chemical biology, 2011. **15**(1): p. 11-21.

106. Naesby, M., et al., *Yeast artificial chromosomes employed for random assembly of biosynthetic pathways and production of diverse compounds in Saccharomyces cerevisiae.* Microbial Cell Factories, 2009. **8**(1): p. 45.

107. Larsson, S., et al., *The generation of fermentation inhibitors during dilute acid hydrolysis of softwood.* Enzyme and Microbial Technology, 1999. **24**(3-4): p. 151-159.

108. Viikari, L., J. Vehmaanpera, and A. Koivula, *Lignocellulosic ethanol: From science to industry.* Biomass & Bioenergy, 2012. **46**: p. 13-24.

109. O'Sullivan, A.C., *Cellulose: The structure slowly unravels.* Cellulose, 1997. **4**(3): p. 173-207.

110. Klinke, H.B., A.B. Thomsen, and B.K. Ahring, *Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass.* Applied Microbiology and Biotechnology, 2004. **66**(1): p. 10-26.

111. Li, F.C., et al., *Arabinose substitution degree in xylan positively affects lignocellulose enzymatic digestibility after various NaOH/H2SO4 pretreatments in Miscanthus.* Bioresource Technology, 2013. **130**: p. 629-637.

112. Dien, B.S., et al., *Screening for L-arabinose fermenting yeasts.* Applied Biochemistry and Biotechnology, 1996. **57-8**: p. 233-242.

113. Santos, R., et al., *Candida queiroziae sp nov., a cellobiose-fermenting yeast species isolated from rotting wood in Atlantic Rain Forest.* Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology, 2011. **99**(3): p. 635-642.

114. Chase, M.W. and M.F. Fay, *Barcoding of Plants and Fungi.* Science, 2009. **325**(5941): p. 682-683.

115. Bellemain, E., et al., *ITS as an environmental DNA barcode for fungi: an in silico approach reveals potential PCR biases.* BMC Microbiology, 2010. **10**: p. 189-189.

116. Heath, E.C., et al., *PENTOSE FERMENTATION BY LACTOBACILLUS PLANTARUM: II. l-ARABINOSE ISOMERASE.* Journal of Biological Chemistry, 1958. **231**(2): p. 1031-1037.

117. Ryabova, O.B., O.M. Chmil, and A.A. Sibirny, *Xylose and cellobiose fermentation to ethanol by the thermotolerant methylotrophic yeast Hansenula polymorpha.* FEMS Yeast Research, 2003. **4**(2): p. 157-164.

118. Galafassi, S., et al., *Lipid production for second generation biodiesel by the oleaginous yeast Rhodotorula graminis.* Bioresource Technology, 2012. **111**: p. 398-403.

119. Kim, S. and B.E. Dale, *Global potential bioethanol production from wasted crops and crop residues.* Biomass and Bioenergy, 2004. **26**(4): p. 361-375.

120. Malins, C., et al., *Wasted: Europe's untapped resource: an assessment of advanced biofuels from wastes and residues.* 2014.

121. Kootstra, A.M.J., et al., *Optimization of the dilute maleic acid pretreatment of wheat straw.* Biotechnology for Biofuels, 2009. **2**(1): p. 31.

122. Menon, V. and M. Rao, *Trends in bioconversion of lignocellulose: Biofuels, platform chemicals & biorefinery concept.* Progress in Energy and Combustion Science, 2012. **38**(4): p. 522-550.

123. Mood, S.H., et al., *Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment.* Renewable and Sustainable Energy Reviews, 2013. **27**: p. 77-93.

124. Chandel, A.K., S.S. da Silva, and O.V. Singh, *Detoxification of Lignocellulose Hydrolysates: Biochemical and Metabolic Engineering Toward White Biotechnology.* BioEnergy Research, 2013. **6**(1): p. 388-401.

125. Chandel, A.K., et al., *Sugarcane bagasse and leaves: Foreseeable biomass of biofuel and bio-products.* Journal of Chemical Technology and Biotechnology, 2012. **87**(1): p. 11-20.

126. Cerveró, J.M., et al., *Enzymatic hydrolysis and fermentation of palm kernel press cake for production of bioethanol.* Enzyme and Microbial Technology, 2010. **46**(3–4): p. 177-184.

127. Whiffin, F., F. Santomauro, and C.J. Chuck, *Toward a microbial palm oil substitute: oleaginous yeasts cultured on lignocellulose.* Biofuels, Bioproducts and Biorefining, 2016. **10**(3): p. 316-334.

128. Fujita, S.-I., et al., *Multiplex PCR Using Internal Transcribed Spacer 1 and 2 Regions for Rapid Detection and Identification of Yeast Strains.* Journal of Clinical Microbiology, 2001. **39**(10): p. 3617-3622.

129. Manter, D.K. and J.M. Vivanco, *Use of the ITS primers, ITS1F and ITS4, to characterize fungal abundance and diversity in mixed-template samples by qPCR and length heterogeneity analysis.* Journal of Microbiological Methods, 2007. **71**(1): p. 7-14.

130. Chen, Y.C., et al., *Identification of Medically Important Yeasts Using PCR-Based Detection of DNA Sequence Polymorphisms in the Internal Transcribed Spacer 2 Region of the rRNA Genes.* Journal of Clinical Microbiology, 2000. **38**(6): p. 2302-2310.

131. Robinson, A., *Genomics–the future of healthcare and medicine.* Prescriber, 2016. **27**(4): p. 51-55.

132. Fleischmann, R.D., et al., *Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.* science, 1995. **269**(5223): p. 496-512.

133. Anderson, S., *Shotgun DNA sequencing using cloned DNase I-generated fragments.* Nucleic Acids Research, 1981. **9**(13): p. 3015-3027.

134. Goffeau, A., et al., *Life with 6000 genes.* Science, 1996. **274**(5287): p. 546-567.

135. International Human Genome Sequencing, C., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**: p. 860.

136. The, C.e.S.C., *Genome Sequence of the Nematode C. elegans: A Platform for Investigating Biology.* Science, 1998. **282**(5396): p. 2012-2018.

137. Adams, M.D., et al., *The genome sequence of Drosophila melanogaster.* Science, 2000. **287**(5461): p. 2185-2195.

138. Kaul, S., et al., *Analysis of the genome sequence of the flowering plant Arabidopsis thaliana.* nature, 2000. **408**(6814): p. 796-815.

139. Grigoriev, I.V., et al., *MycoCosm portal: gearing up for 1000 fungal genomes.* Nucleic Acids Research, 2014. **42**(D1): p. D699-D704.

140. Galagan, J.E., et al., *Genomics of the fungal kingdom: insights into eukaryotic biology.* Genome research, 2005. **15**(12): p. 1620-1631.

141. Mohanta, T.K. and H. Bae, *The diversity of fungal genome.* Biol Proced Online, 2015. **17**: p. 8.

142. Dujon, B., *The yeast genome project: what did we learn?* Trends in Genetics, 1996. **12**(7): p. 263-270.

143. Cherry, J.M., et al., *Saccharomyces Genome Database: the genomics resource of budding yeast.* Nucleic Acids Research, 2012. **40**(D1): p. D700-D705.

144. Dujon, B., et al., *Genome evolution in yeasts.* Nature, 2004. **430**: p. 35.

145. Kellis, M., et al., *Sequencing and comparison of yeast species to identify genes and regulatory elements.* Nature, 2003. **423**: p. 241.

146. Suh, S.-O., et al., *Phylogenetics of Saccharomycetales, the ascomycete yeasts.* Mycologia, 2006. **98**(6): p. 1006-1017.

147. Spadaro, D., et al., *A new strain of Metschnikowia fructicola for postharvest control of Penicillium expansum and patulin accumulation on four cultivars of apple.* Postharvest Biology and Technology, 2013. **75**: p. 1-8.

148. Kluyver, A., J. Van der Walt, and A. Van Triet, *Pulcherrimin, the pigment of Candida pulcherrima.* Proceedings of the National Academy of Sciences, 1953. **39**(7): p. 583-593.

149. Roberts, C., *The effect of iron and other factors on the production of pigment by the yeast Torulopsis pulcherrima.* American journal of botany, 1946: p. 237-244.

150. Metzker, M.L., *Sequencing technologies — the next generation.* Nature Reviews Genetics, 2009. **11**: p. 31.

151. Batista, T.M., et al., *Draft Genome Sequence of Metschnikowia australis Strain UFMG-CM-Y6158, an Extremophile Marine Yeast Endemic to Antarctica.* Genome Announcements, 2017. **5**(20): p. e00328-17.

152. Storck, R., *Nucleotide Composition of Nucleic Acids of Fungi II. Deoxyribonucleic Acids.* Journal of Bacteriology, 1966. **91**(1): p. 227-230.

153. Riley, R., et al., *Comparative genomics of biotechnologically important yeasts.* Proceedings of the National Academy of Sciences, 2016. **113**(35): p. 9882-9887.

154. Kawaguchi, Y., et al., *The codon CUG is read as serine in an asporogenic yeast Candida cylindracea.* Nature, 1989. **341**(6238): p. 164-166.

155. Santos, M.A., et al., *The genetic code of the fungal CTG clade.* Comptes rendus biologies, 2011. **334**(8): p. 607-611.

156. Gomes, A.C., et al., *A genetic code alteration generates a proteome of high diversity in the human pathogen Candida albicans.* Genome Biology, 2007. **8**(10): p. R206.

157. Sabirova, J.S., et al., *The 'LipoYeasts' project: using the oleaginous yeast Yarrowia lipolytica in combination with specific bacterial genes for the bioconversion of lipids, fats and oils into high‐value products.* Microbial biotechnology, 2011. **4**(1): p. 47-54.

158. Ashburner, M., et al., *Gene Ontology: tool for the unification of biology.* Nature genetics, 2000. **25**(1): p. 25-29.

159. Shen, Q., et al., *Comparative genome analysis of the oleaginous yeast Trichosporon fermentans reveals its potential applications in lipid accumulation.* Microbiological Research, 2016. **192**(Supplement C): p. 203-210.

160. Chatzifragkou, A., et al., *Commercial sugars as substrates for lipid accumulation in Cunninghamella echinulata and Mortierella isabellina fungi.* European Journal of Lipid Science and Technology, 2010. **112**(9): p. 1048-1057.

161. Santamauro, F., et al., *Low-cost lipid production by oleaginous yeast cultured in non-sterile conditions using waste resources.* Biotechnol Biofuel, 2014. **7**.

162. Lavarack, B.P., G.J. Griffin, and D. Rodman, *The acid hydrolysis of sugarcane bagasse hemicellulose to produce xylose, arabinose, glucose and other products.* Biomass and Bioenergy, 2002. **23**(5): p. 367-380.

163. Palmqvist, E., et al., *Main and interaction effects of acetic acid, furfural, and p-hydroxybenzoic acid on growth and ethanol productivity of yeasts.* Biotechnology and Bioengineering, 1999. **63**(1): p. 46-55.

164. Pan, L.X., et al., *Isolation of the oleaginous yeasts from the soil and studies of their lipid-producing capacities.* Food Technology and Biotechnology, 2009. **47**(2): p. 215-220.

165. Freedman, B., R.O. Butterfield, and E.H. Pryde, *Transesterification kinetics of soybean oil 1.* Journal of the American Oil Chemists' Society, 1986. **63**(10): p. 1375-1380.

166. Sambrook, J. and D.W. Russell, *Purification of Nucleic Acids by Extraction with Phenol:Chloroform.* Cold Spring Harbor Protocols, 2006. **2006**(1): p. pdb.prot4455.

167. Chang, J.-Y., *A Two-Stage Mechanism for the Reductive Unfolding of Disulfide-containing Proteins.* Journal of Biological Chemistry, 1997. **272**(1): p. 69-75.

168. Green, M.R. and J. Sambrook, *Isolation of High-Molecular-Weight DNA Using Organic Solvents.* Cold Spring Harbor Protocols, 2017. **2017**(4): p. pdb.prot093450.